Mathematical Modelling of DNA Replication

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

d o c t o r r e r u m n a t u r a l i u m
(Dr. rer. nat.)
im Fach Biophysik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
der Humboldt-Universität zu Berlin

von
Diplom-Physikerin Anneke Brümmer
20.10.1977 in Hamburg

Präsident der der Humboldt-Universität zu Berlin:
Prof. Dr. Dr. h.c. Christoph Markschies

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:
Prof. Dr. Lutz-Helmut Schön

Gutachter:
1. Prof. Dr. Peter Hammerstein
2. Prof. Thomas Höfer, PhD
3. Prof. Dr. Lilia Alberghina

Tag der mündlichen Prüfung: 08.07.2010

Abstract

Before a cell divides into two daughter cells, its entire genetic material has to be copied without errors and exactly once. In eukaryotic cells, a vast amount of replication origins exists that enable the replication of the DNA to initiate simultaneously from many origins in parallel, thereby contributing to a relatively rapid duplication of the genome. The initiation of DNA replication from the replication origins is a tightly controlled process. The molecular machinery involved in this process in budding yeast has been identified in the past decades, but questions remain concerning their precise dynamical behavior and interactions.

In order to restrict the initiation of DNA replication to once per cell cycle, the activation of the origins proceeds in two temporally separated phases, the licensing and the firing phase. A part of the replication machinery, including the presumptive DNA helicase, is assembled at the origins in the first phase, and is completed by loading of DNA polymerases in the second phase. The temporal separation of origin licensing and firing is tightly regulated by the activity of cyclin-dependent kinases (Cdks).

In this work, a mathematical model for DNA replication in budding yeast is provided. Based on a multitude of experimental studies, a molecular interaction network is constructed and translated into balance equations for all molecule complexes assembled at the replication origins, free molecule complexes and all phosphorylation states. Initial protein concentrations could be taken from measurements. The kinetic parameters of the mathematical model are determined by using an optimization approach. Firstly, the biological functionality of the system is defined by means of four functional systems properties, the fraction of activated origins, the number of rereplicating origins and the rate of origin activation, measured by its mean time and duration. Secondly, the biological functionality of the entire system is maximized as a function of its kinetic parameters.

The parameterized model accounts for the experimentally observed distribution of activation times of early replication origins and at the same time realizes the strict inhibition of DNA rereplication. Analysis of the kinetics of origin firing revealed that the prevention of DNA rereplication relies on a time delay between the licensing and firing of replication origins, which, however, limits the rate of origin activation. The multisite phosphorylation of two target proteins of the S phase cyclin-dependent kinase (S-Cdk), Sld2 and Sld3, is essential for creating a robust time delay before the activation of replication origins and at the same time in contributing to a synchronous initiation of DNA replication at several replication origins. The mathematical model rationalizes experimentally realized deregulations in the activity of Cdks and quantifies the resulting disorders in the kinetics of origin activation.

Furthermore, the kinetics of origin activation calculated with the mathematical model is utilized to predict the consequences of specific deregulations in the activation of replication origins on the process of DNA replication during S phase, which is quantified by the duration of the DNA synthesis period and the distribution of DNA replicon sizes.

In summary, a consistent model for DNA replication in budding yeast is developed and analyzed extensively. The connection of the initiation kinetics of DNA replication and its dynamics during S phase of the cell cycle, allows to comprehensively study the potential sources of chromosomal rearrangements.

Zusammenfassung

Bevor sich eine Zelle in zwei Tochterzellen teilt, muß ihr genetisches Material fehlerfrei und genau einmal kopiert werden. In eukaryotischen Zellen existieren eine große Anzahl von Replikationsstartpunkten, die sogenannten Replikationsorigins. Diese ermöglichen es, daß die DNA Replikation von mehreren Origins aus zeitgleich gestartet wird, was zu einer relativ kurzen Verdoppelungzeit, auch von großen Genomen, beiträgt. Die Initiierung der DNA Replikation an den Replikationsorigins ist ein stark kontrollierter Prozess. Die molekulare Maschinerie, die an diesem Prozess in Hefezellen beteiligt ist, wurde innerhalb der letzten Jahrzehnte identifiziert. Es gibt jedoch offene Fragen bezüglich ihres genauen dynamischen Verhaltens und Zusammenwirkens.

Um die Initiierung der DNA Replikation auf exakt einmal pro Zellzyklus zu beschränken, erfolgt die Aktivierung der Replikationsorigins in zwei zeitlich getrennten Phasen, der sogenannten Lizensierungsphase und der Feuerphase. Ein Teil der Replikationsmaschinerie, einschließlich der mutmaßlichen DNA Helikase, wird in der ersten Phase an den Origins zusammengestellt, und in der zweiten Phase durch die Bindung der DNA Polymerasen komplettiert. Die zeitliche Trennung von Originlizensierungsphase und Originfeuerphase wird durch die Aktivität von Cyclin-abhängigen Kinasen (Cdk) strikt reguliert.

In dieser Arbeit wird ein mathematisches Modell für die DNA Replikation in Hefezellen präsentiert. Basierend auf einer Vielzahl von experimentellen Studien, wird ein molekulares Interaktionsnetzwerk konstruiert, und als ein System von Balancegleichungen für alle Molekülkomplexe an den Replikationsorigins, alle freien Molekülkomplexe und alle Phosphorylierungszustände mathematisch formuliert. Die Anfangsbedingungen für die Proteinkonzentrationen wurden experimentellen Messungen entnommen. Die kinetischen Parameter des mathematische Models werden mit Hilfe eines Optimierungsansatzes bestimmt. Zunächst wird die biologische Funktionalität des Systems mit Hilfe von vier funktionellen Eigenschaften definiert, der Anteil der aktivierten Origins, der Anzahl der rereplizierten Origins, und der Rate der Originaktivierung, welche durch ihren mittleren Zeitpunkt und ihre Dauer bestimmt sind. Anschließend wird die biologische Funktionalität des gesamten Systems als Funktion der kinetischen Parameter maximiert.

Das parameterisierte Modell stimmt mit den experimentell beobachteten Verteilungen der Aktivierungszeiten der frühen Replikationsorigins überein und realisiert gleichzeitig die strikte Inhibierung der DNA Rereplikation. Die Analyse der Originaktivierungskinetiken verdeutlicht, daß die Verhinderung der DNA Rereplikation auf der zeitlichen Trennung der Lizensierung und des Feuerns der Replikationsorigins beruht, welche aber auch die Geschwindigkeit der Originaktivierung limitiert. Die multiple Phosphorylierung zweier Targetproteine, Sld2 und Sld3, der S Phasen-Cyclin-abhängigen Kinase, S-Cdk, stellt sich als essentiell für die Schaffung eines robusten Zeitverzögerungsmoduls vor der Aktivierung der Replikationsorigins heraus und trägt gleichzeitig zu einer möglichst synchronen Initiierung der DNA Replikation an den verschiedenen Replikationsorigins bei. Das mathematische Model gibt die experimentell realisierte Deregulierung der Cdk-Aktivität wieder und quantifiziert die resultierende, gestörte Originaktivierungskinetik.

Desweiteren wird die Originaktivierungskinetik, die mit dem mathematischen Modell berechnet werden kann, genutzt, um die Konsequenzen einer spezifischen Deregulierung in der Initiierung der DNA Replikation auf den Prozess der DNA Replikation in der S Phase vorherzusagen. Der DNA Replikationsprozess wird durch die Dauer der DNA Synthese und die Verteilung der DNA Replikongrößen charakterisiert.

Zusammenfassend wird ein realistisches Modell für die DNA Replikation in Hefezellen

entwickelt und umfangreich analysiert. Die Verbindung der Aktivierungskinetik der Replikationsorigins mit der Dynamik der DNA Replikation in der S Phase des Zellzyklus, ermöglicht es potentielle Quellen von chromosomalen Umordnungen umfassend zu untersuchen.

Contents

1 Introduction

1.1 DNA replication in budding yeast

The proliferation of a cell population is driven by the cell cycles of the individual cells, which, while growing, duplicate all their components and divide into two daughter cells.
The eukaryotic cell cycle is divided into four phases, the mitotic phase (M phase), the synthesis phase (S phase) and two gap phases in between (G1 phase and G2 phase). After the exit from the previous mitosis, the new cell cycle starts with the G1 phase, in which the cell increases in size and prepares for the upcoming DNA replication, occurring during the S phase. After the completed DNA synthesis, in the G2 phase, the cell gets ready for the forthcoming separation of sister chromatids, which are then segregated to the two daughter cells at the cell division during the M phase. All phase transitions during the cell cycle are tightly controlled by the activity of cyclin dependent kinases (Cdks), whose concentration is low during G1 phase and rises from S to M phase (Figure 1.1).

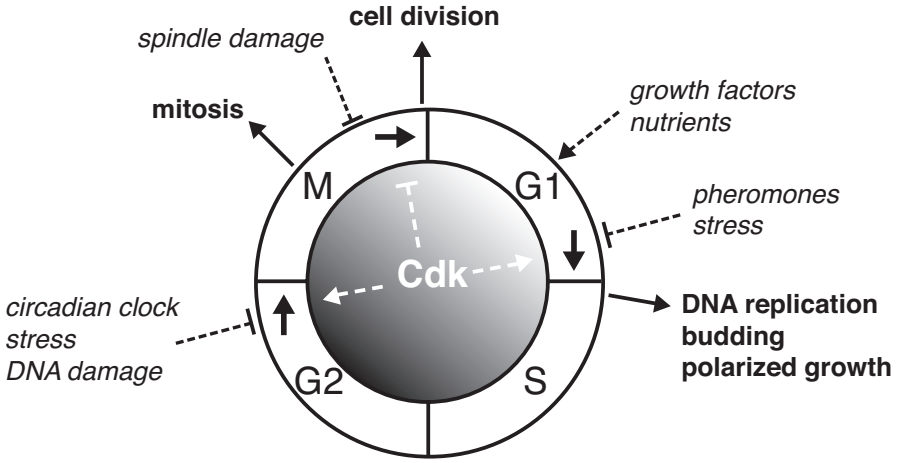


Figure 1.1: The progression of the cell cycle is regulated by the activity of Cdks (white dashed lines; activating (arrows) and inhibiting (bar) regulation). The different phases of the cell cycle are also influenced by other cellular and external signals (black dashed lines; arrows: activating and bars: inhibiting) and specific processes in the duplication of the cell are initiated (black arrows). Figure redrawn from [18].

The replication of the genome is a critical process in the duplication of cells. The genetic material has to be copied without errors and exactly once per cell cycle. In bacteria, DNA replication is initiated from a single origin, whereas eukaryotes use hundreds to thousands of origins

in parallel to accelerate the doubling of the long genomes [90]. In the genome of budding yeast, *Saccharomyces Cerevisiae*, more than 500 likely autonomously replicating sequences (ARSs) are identified [60], which serve as potential origins for DNA replication. Not all origins are activated in every cell cycle, so that the composition as well as the quantity of origins, from which DNA replication is finally initiated in a cell cycle, varies. While the potential origins are present in great excess in fission yeast and are consequently activated with low efficiency in a cell cycle [87], the efficiency of origin activation in budding yeast is generally higher, but still the factors that determine the different efficiencies of the origins are not fully understood [67, 90].

The ARSs are marked with an origin recognition complex (ORC) throughout the cell cycle (Figure 1.2), which provides a platform for the assembly of the replication complex (RC). The formation of the RCs at the replication origins proceeds in a defined sequence of binding and phosphorylation events occurring in two consecutive phases, origin licensing and firing, which starts in the G1 phase and ends with the onset of DNA synthesis in S phase [22].

During the licensing phase, the putative DNA helicase, Mcm2-7, is loaded to the replication origins in an inactive form, building the pre-replication complexes (pre-RCs). The activation of the G1 phase cyclin-dependent kinase G1-Cdk, which consists of the regulatory cyclins Cln1 or Cln2 and the enzymatically active kinase Cdc28, leads to the inhibition of origin licensing. Simultaneously, G1-Cdk triggers the activation of another, S phase cyclin-dependent kinase, S-Cdk, composed of the regulatory cyclins Clb5 or Clb6 and the kinase Cdc28. S-Cdk together with another kinase, Ddk (consisting of the regulatory subunit Dbf4 and the kinase subunit Cdc7) then, in the firing phase, phosphorylate components of Mcm2-7 that are part of the pre-RCs. This enables the binding of the essential Cdc45 to the origins and the formation of pre-initiation complexes (pre-ICs). S-Cdk also phosphorylates other important molecules, Sld2 and Sld3, that facilitate the loading of GINS and the DNA polymerase and the formation of the RCs at the replication origins and, thus, the initiation of DNA synthesis (Figure 1.2).

The temporal separation of the licensing and the firing phase, which is tightly controlled by the activity of Cdks, ensures that no origin becomes reactivated within the same cell cycle. When the origins are activated by Cdks in the firing phase, licensing is inhibited, by the same enzymes, preventing the further formation of pre-RCs at the origins and the rereplication of DNA. But, occasionally, this mechanism causes some origins to be inhibited prematurely, before being licensed and activated once, or some origins that fail to complete their pre-RC, before the inhibition of licensing reactions sets in. These origins remain inactive during S phase and are called silent origins [90].

During S phase, the replication machinery containing the DNA helicase and DNA polymerases starts the synthesis of new DNA bidirectionally from the replication origin, as soon as its RC has formed.

The activation of the origins is temporally coordinated, such that a part of the origins, known as early origins, initiates the replication of more than half of the budding yeast genome ($\sim$ 7-8 Mb) in early S phase. Another part of the origins, called late origins, becomes active later in S phase and completes the duplication of the genome. Early and late origins are licensed at the same time, while the activation of late origins is temporally delayed [71, 35]. A relationship probably exists between the accessibility of the chromatin, containing the origin, and its time of replication. However, a correlation with the activity of gene transcription, as observed in other organisms, has not been confirmed for budding yeast [28]. The specific biochemical factors

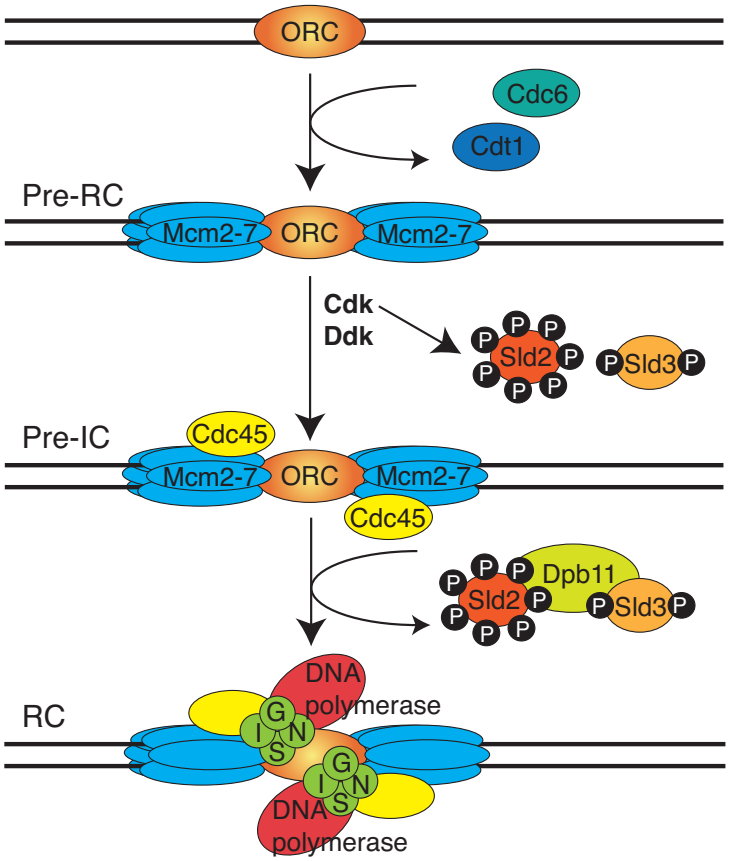


Figure 1.2: The formation of replication complexes (RC) at the replication origins proceeds in a defined sequence. Pre-replication complexes (pre-RC) containing the presumptive DNA helicase, Mcm2-7, are assembled during the licensing phase and pre-initiation complexes (pre-IC) are formed following the activation of S-Cdk. At the same time, S-Cdk phosphorylates Sld2 and Sld3 that finally enable the loading of DNA polymerases completing the replication complex (RC).

that regulate the differences between early and late activated origins and lead to the temporal regulation of the DNA replication within the S phase, are unknown [90].

A wide difference in the activation times of neighboring replication origins might result in another fraction of origins that remains inactive during S phase. These origins, which are called dormant origins [90], become replicated passively by their neighboring replication machinery before they would have been activated themselves.

The transition of the G1 to S phase is an important step in the cell cycle that must be tightly controlled. Deregulations in the G1/S transition may strongly affect the process of DNA synthesis during S phase. The incomplete duplication of the genome causes a loss of the genetic material. DNA rereplication may result in fragmented and partially doubled daughter chromo-

somes and, thus, contribute to the emergence of genomic instability [73, 31]. Indeed, most genes that are mutated in human cancers are influencing the G1 to S phase transition.

1.2 Related mathematical models

To rationalize the kinetic behavior of a molecular interaction network governing a biological process, mathematical modeling is very valuable. The identification and analysis of system level properties that only emerge from the dynamic interaction of these molecules contribute to a quantitative understanding of the underlying biological process.

Mathematical modeling of the cell cycle regulation has a long tradition and has been successfully applied to study biological questions. The development of mathematical models started around 50 years ago with the quantitative analysis of the influence of cell growth on cell division, which was experimentally identified by Prescott [66, 18]. Further experimental discoveries of molecular interactions involved in the control of the cell cycle, contributed to the development of a variety of models focussed on the molecular regulation of the cell cycle and specific transitions within the cell cycle.

Deterministic models based on ordinary differential equations (ODEs) have been employed to describe the kinetics of cyclin-dependent kinases and its regulators, such as inhibitors or transcription factors, to analyze the cell cycle control in budding yeast [12, 13] and fission yeast [63]. A very simplified model was reported generating an oscillating cell cycle in of the eukaryotic cell cycle by using delay differential equations (DDE) [76].

Several kinetic models focused on specific transitions within the cell cycle and have been successful in quantifying particular molecular control mechanisms. Recently, a positive feedback mechanism, acting on the level of gene expression of the cyclins Cln1 and Cln2 was proposed to induce a robust and switch-like entry into the G1 phase of the cell cycle [74]. Likewise, the proteolysis of cyclins was shown to be insufficient for the unidirectionality of cell cycle transitions after a longer time period, specifically for the mitotic exit. Instead, the presence of an additional feedback loop was proposed, which ensures the irreversibility of the mitotic exit in the long run [51]. Also, the regulation of the G1 to S phase transition was analyzed in a kinetic model focussing on the impact of the cell size on this transition [6].

Mathematical models were also used to evaluate the DNA replication dynamics during S phase. Generally, stochastic models are used to describe the DNA replication dynamics on a phenomenological level, mostly without connection to the molecular processes controlling the activation of the individual origins. A universal model for the dynamics of DNA replication in eukaryotes is reported that bases on experimentally determined origin activation profiles [29, 30]. A stochastic hybrid model for the DNA replication dynamics in fission yeast is proposed that couples the discrete transitions between the states of an origin to the continuous movement of the replication machinery along the DNA. This model is predicated upon experimental information about the location of each putative origin and its measured activation propensity. To explain the experimentally measured duration of the S phase in fission yeast, the model proposes the existence and redistribution of a limiting factor during the activation of replication origins [53]. Another model for the DNA replication dynamics in budding yeast uses the acti-

vation times of replication origins and their positions on the chromosomes from measured DNA replication profiles to simulate the replication dynamics under normal conditions and of distinct perturbed conditions [75]. A few models exist that apply methods from theoretical statistical physics to the DNA replication dynamics [55, 40, 39].

Finally, mathematical models have also been reported for other dynamic processes on the DNA molecule in eukaryotic cells, like the assembly of the molecular machinery performing nucleotide excision repair (NER) of single strand breaks in mammalian DNA [65, 52] or the initiation of single-gene transcription in human cells, which includes a molecular complex formation on the DNA and specific structural rearrangements of the DNA strand that finally lead to transcriptional cycles on the cell population level [19].

1.3 Research objectives

A full cell cycle in budding yeast takes around 2 hours [21]. Within this short time period, the complete genome has to be replicated, separated into two identical copies, and distributed to mother and daughter cells. The duplication of the genome would proceed most efficiently, if all replication origins are initiated simultaneously and the DNA is replicated bidirectionally from all origins in parallel. However, a temporal distribution of activation times of replication origins is observed [67, 88], with differences between early and late origins as well as variations even among the early origins. The reasons for these characteristic variations are unclear, although some factors, such as an effect of the chromatin environment of the late replication origin have been observed [24].

A variety of experimental studies have focussed on the molecular processes leading to the initiation of DNA replication in budding yeast, but up to now a quantitative understanding of the molecular network that leads to the activation of replication origins is lacking.

In Chapter 2 of this work, a molecular network model for the initiation of DNA replication is constructed and translated into a mathematical model to elucidate the regulatory mechanisms by which a large number of origins is activated near-simultaneous, while exhibiting the physiologically required functionality. The molecular interaction network governing the initiation of DNA replication is constructed making use of the large amount of experimental studies in budding yeast. The concentrations of the participating molecules are taken from measured protein expression data [27], while most kinetic parameters describing molecular reactions, such as binding, dissociation, phosphorylation, dephosphorylation or degradation reactions, have not been (and, in many cases, cannot be) determined experimentally. To identify parameter sets that reproduce a realistic origin activation kinetics, which has indeed been measured [67, 88], the different parameter types are first constrained to biochemically allowed ranges and then optimized to generate a functional kinetics of origin activation. The functionality of the kinetics of origin activation, that is the systems performance, is measured by means of four functional systems properties, the number of activated replication origins, the number of rereplicating origins, and the rate of origin activation, defined by the mean value and the standard deviation of the distribution of origin activation times.

The derived parametrized mathematical model is used, in Chapter 3, to reproduce and quantify

experimentally realized deregulations in the molecular interaction network, thereby validating the mathematical model.

Specific questions addressed by the mathematical model are:

- Are there constraints in the functionality of the kinetics of origin activation that can be identified by interrelations between functional systems properties?
- Are the individual functional systems properties controlled by specific molecule concentrations and kinetic rate constants?
- What is the role of the multiple phosphorylation of several molecules components by Cdks in the regulation of DNA replication initiation?

The consequences of a particular kinetics of origin activation on the DNA replication dynamics during S phase are addressed in a second mathematical model in Chapter 4. This subsequent model uses the simulated kinetics of origin activation and connects it to the dynamics of DNA replication, allowing to quantify and evaluate the impact of deregulations in the G1 to S phase transition on the process of DNA synthesis during S phase.

2 Kinetic model for the initiation of DNA replication

In this chapter, a molecular network for the regulation of the initiation of DNA replication in budding yeast is constructed and translated into a mathematical model. The biological functionality of the molecular regulatory system is defined. Kinetic parameterizations of the mathematical model are generated systematically by maximizing the functionality of the biological system as a function of its kinetic parameters.

2.1 Molecular regulatory network for the initiation of DNA replication

Based on a vast amount of experimental data, we constructed a molecular network for the regulation of DNA replication initiation in budding yeast (Figure 2.1). The network considers the essential molecules and reactions known to be involved in the assembly of the RC at the early origins. A sequence of reversible binding reactions and directed de-/phosphorylations, denoted as transitions of the origin between the states S0 to S9 (Table 2.1), leads to the final formation of the RC at the origin (in state S9) and the initiation of DNA replication from that origin.

In the following, the reactions of the molecular network (Figure 2.1) are described in detail, in the order of three temporally consecutive phases: licensing phase, intermediate phase, and firing phase. The reactions involved in mechanisms responsible for the inhibition of DNA rereplication are specified separately.

2.1.1 Licensing phase

During the licensing phase (blue in Figure 2.1), which starts after the exit from mitosis, Mcm2-7 proteins, the putative replicative helicase complex, are loaded to the origins in a highly regulated process.

The starting positions for DNA replication, the ARS, are marked with ORCs, which consist of six subunits, Orc1 to Orc6, that stay attached to the DNA throughout the cell cycle [48]. The Mcm2-7 complex, which comprises six subunits, Mcm2 to Mcm7, forms a complex with Cdt1 (reactions 4 and 5) in the cytoplasm [78]. A nuclear localization signal (NLS) in Mcm2-7 leads to the import of the Cdt1-Mcm2-7 complex into the nucleus [49]. Two ATPases, ORC and Cdc6, are already bound to the origins (reaction 1), before Cdt1-Mcm2-7 associates with these molecules at the DNA (reaction 3). ATP hydrolysis by Cdc6 then allows a tight connection of origin DNA and Mcm2-7 proteins. Subsequently, Cdt1 is released from the origins and

Table 2.1: The states S0 to S8 represent temporary molecular complexes in the assembly of the replication complex, state S9 in the kinetic model. Certain molecular complexes are referred to by synonyms.

State	Molecular complex	Synonym
S0	ORC	
S1	ORC-Cdc6	
S2	ORC-Cdc6-Cdt1-Mcm2-7	
S3	ORC-Mcm2-7	pre-replication complex (pre-RC)
S4	ORC-Mcm2-7_{p}	
S5	ORC-Mcm2-7_{pp}	
S6	ORC-Mcm2-7_{pp}-Cdc45	pre-initiation complex (pre-IC)
S7	ORC-Mcm2-7_{pp}-Cdc45-11-3-2	
S8	ORC-Mcm2-7_{pp}-Cdc45-11-3-2-GINS/polymerase	
S9	ORC-Mcm2-7_{pp}-Cdc45-GINS-polymerase	replication complex (RC)

Cdc6 association becomes destabilized (reaction 9). ATP hydrolysis by ORC then completes the Mcm2-7 loading reaction [68], leaving tightly bound pre-RCs at the replication origins (state S3).

2.1.2 Intermediate period

With the activation of the G1-phase cyclin-dependent kinase, G1-Cdk, which occurs after the completion of pre-RCs at all replication origins under normal conditions, the next step in the initiation of DNA replication commences. G1-Cdk causes the inhibition of origin licensing and, at the same time, the activation of the S-phase cyclin-dependent kinase, S-Cdk.

Phosphorylation of Cdc6 by G1-Cdk causes a rapid degradation of Cdc6p (reaction 2) via an SCF^{Cdc4}-mediated ubiquitination and proteolysis [23]. Phosphorylation of Mcm2-7 by G1-Cdk (reaction 6) blocks the NLS and causes the activation of an adjacent nuclear export signal (NES) in the Mcm2-7 molecule, which leads to the export of free Mcm2-7 or Mcm2-7 in complex with Cdt1 (reaction 8) from the nucleus [49]. Cdt1 is also exported from the nucleus (reaction 7), but it is not clear, whether this process is dependent on a phosphorylation by G1-Cdk [79].

Sic1 is a stoichiometric inhibitor of S-Cdk, that binds to S-Cdk (reaction 12) and inhibits its activity during the licensing phase. Sic1 is phosphorylated by G1-Cdk at up to nine serine and threonine residues in a random sequence (reaction 10). Phosphorylations can be reversed by a phosphatase, presumably Cdc14 [83, 7], that is also assumed to act in a random sequence (reaction 11). Any combination of at least six phosphorylations in Sic1 leads to the recognition of the Sic1-S-Cdk complex by Cdc4 and an SCF-mediated ubiquitination and subsequent proteolysis of Sic1 (reaction 13) [58, 20]. Degradation of Sic1 then results in a delayed release of active S-Cdk at late G1 phase.

2.1.3 Firing phase

S-Cdk promotes the final phase in the preparation for DNA replication via phosphorylation of several target proteins.

Two proteins, Sld2 and Sld3, are phosphorylated at multiple residues by S-Cdk before they can bind to Dpb11 to form the Sld3-Dpb11-Sld2 (11-3-2) complex, which associates with the origins only transiently and is supposed to catalyze the recruitment of the GINS complex together with DNA polymerases. Sld2 is phosphorylated by S-Cdk randomly at six serine and threonine residues (reaction 14), causing a conformational change in the protein that exposes the essential threonine 84 residue for phosphorylation by S-Cdk (reaction 16), being the docking site for the binding to Dpb11 (reaction 18) [54, 77]. Phosphorylation of two residues, one serine and one threonine, in Sld3 by S-Cdk (reaction 14) is essential for the binding to Dpb11 (reaction 19) [89, 81]. Phosphorylations in both proteins can be reversed by a phosphatase, presumably Cdc14 [7], which is also assumed to act in a random sequence (reactions 15 and 17). It is not clear, whether Sld2p, Sld3p and Dpb11 assemble in a preferred sequence or randomly to form the 11-3-2 complex [80].

Mcm2-7 proteins bound to the origin DNA are phosphorylated by two different kinases. S-Cdk is supposed to act as a priming kinase (reaction 20) [62] for the multiple phosphorylation of Mcm2-7 by Ddk (reaction 21) [72]. This multisite phosphorylation is assumed to be processive and allows the tight binding of the essential molecule Cdc45 to the origins (reaction 22) [41], forming the pre-IC at the origin. The 11-3-2 complex then binds to the pre-IC (reaction 23) and catalyzes the final step in the formation of the replication complexes, the loading of the GINS complex together with the DNA polymerase (reaction 24). The 11-3-2 complex associates only transiently with the origins and becomes destabilized (reaction 25) after the completion of the RC. The free 11-3-2 complex is then able to bind to the pre-RC at another origin and again catalyze the DNA polymerase loading reaction at that origins.

RCs are assumed to become destabilized after the complete synthesis of their respective DNA replicons and, subsequently, the molecular components of the RC become available for the initiation of further origins (reaction 27). Similarly, origins that have not been inhibited by a phosphorylation at their ORC (explained in detail in the next section) at the time of RC completion are allowed to re-enter the replication initiation network at state S0 (reaction 27), after termination of the synthesis of their respective DNA replicon. These origins may become licensed a second time and initiate DNA rereplication, as long as licensing molecules are available and ORC components remain unphosphorylated.

2.1.4 Prevention of DNA rereplication

Several independent mechanisms prevent the re-initiation of an already fired replication origin [2, 9, 59]. It is not clear, whether these mechanisms are redundant or are needed in combination to ensure the inhibition of DNA rereplication. The basic principle of all these mechanism is the inhibition of licensing reactions, such that the formation of pre-RCs is no longer possible, once the firing phase has started.

Three main mechanisms are considered in the mathematical model that contribute to the tem-

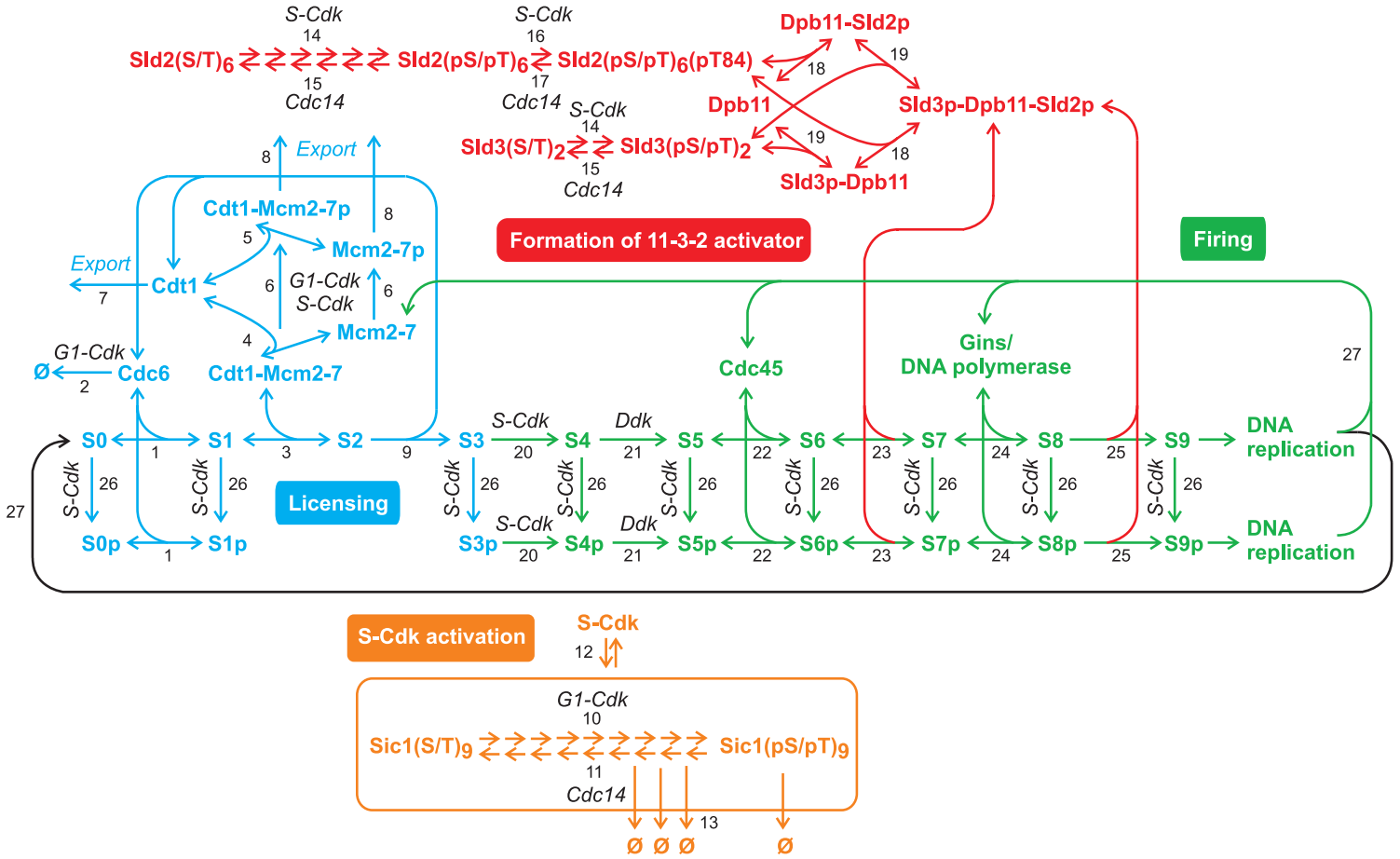


Figure 2.1: The molecular reaction network comprises all essential reactions in the regulation of the initiation of DNA replication at the replication origins. The assembly of the replication complex (S9) occurs via the transient molecular complexes, S0 to S8, and is grouped in four modules, licensing, S-Cdk activation, 11-3-2 activator, and firing module, that proceed in three temporally successive phases: licensing phase (reactions colored in blue), intermediate period (reactions colored in orange) and firing phase (reactions colored in red and green).

poral limitation of the licensing phase. Two mechanisms rely on phosphorylation by G1-Cdk, that are the degradation of Cdc6 after phosphorylation by G1-Cdk (reaction 2) [36] and the export of free Mcm2-7 proteins after phosphorylation by G1-Cdk (reaction 8) [49] that is associated with the export of free Cdt1 from the nucleus (reaction 7). The third mechanism is based on the distinct temporal regulation of the activity of S-Cdk during the licensing and firing phase. S-Cdk phosphorylates ORC components (reaction 26), presumably Orc6 and Orc2 [59], leading to the inhibition of Cdt1 binding to the origins (reaction 3), probably for sterical reasons [59, 14]. In this way the loading of the putative helicase, Mcm2-7, to the origins (transition from state S1 to state S3) becomes impossible. S-Cdk also phosphorylates free Mcm2-7 proteins that are not attached to the DNA (reaction 6) and contributes to the inhibition of origin licensing by blocking the NLS in Mcm2-7 and inducing its export into to cytosol (reaction 8).

Under normal conditions, the mechanisms preventing DNA rereplication do not interfere with the preparation for origin firing and the formation of the RCs. Normally, origin licensing is completed at almost all origins before the firing phase starts, and the formation of new pre-RCs at the origins is already inhibited before the completed RC leaves the origin performing DNA synthesis.

2.2 Mathematical model for the formation of replication complexes at the origins

To analyze the initiation of DNA replication in budding yeast a mathematical model is constructed based on the molecular interaction network depicted in Figure 2.1. The mathematical model is formulated as a system of ODEs neglecting spatial dependencies in the molecule concentrations. This is supposed to appropriately describe the kinetics of DNA replication initiation, since all considered reactions occur within the limited local environment of the cell nucleus and the diffusion of molecules is assumed to be fast. A stochastic approach for the kinetics of replication origin initiation is not required, as the molecule numbers are all large enough (ranging from > 100 to several thousand molecules) and fluctuations in the molecule numbers are supposed to have no effect on the kinetics of origin initiation.

Balance equations are formed for all molecule complexes that are free in the nucleoplasm and assembled at the replication origins and for all phosphorylation states. The rate laws for the elementary reactions, that are binding, dissociation, phosphorylation, dephosphorylation, degradation, and nuclear export reactions, are derived assuming mass-action kinetics. The complete mathematical model consists of a system of 116 ODEs with 14 initial protein concentrations, 51 kinetic rate constants, and a sigmoidal input function for the activation of G1-Cdk.

2.2.1 Variables and initial conditions

The molecule complexes and intermediate phosphorylation states that are considered as variables in the mathematical model are listed in Table 2.2. The corresponding symbolic abbreviation and the initial amount of the molecule used in the ODE system is indicated.

Table 2.2: Variables of the model with corresponding symbols and initial numbers.

Variable	**Symbol**	**Number (Conc.)**
States of replication origins		
Orc	S_0	380 (127 nM)
Orc/Cdc6	S_1	0
Orc/Cdc6/Cdt1/Mcm2-7	S_2	0
Orc/Mcm2-7	S_3	0
Orc/Mcm2-7/S-Cdk	S_{Cdk}	0
Orc/Mcm2-7p	S_4	0
Orc/Mcm2-7p/Ddk	S_{Ddk}	0
Orc/Mcm2-7pp	S_5	0
Orc/Mcm2-7pp/Cdc45	S_6	0
Orc/Mcm2-7pp/Cdc45/11-3-2	S_7	0
Orc/Mcm2-7pp/Cdc45/11-3-2/GINS/DNA polymerase	S_8	0
Orc/Mcm2-7pp/Cdc45/GINS/DNA polymerase	S_9	0
S-Cdk bound Orc in states $i = 0,1,3,4,5,6,7,8,9$	Sc_i	0
Phosphorylated Orc in states $i = 0,1,3,4,5,6,7,8,9$	Sp_i	0
Binding factors		
Cdc6	F_1	700 (233 nM)
Cdt1	F_{2a}	2200 (733 nM)
Mcm2-7	F_{2b}	2000 (667 nM)
Mcm2-7p	F_{2bp}	0
Cdt1–Mcm2-7	F_2	0
Cdt1–Mcm2-7p	F_{2p}	0
Cdc45	F_4	1700 (567 nM)
GINS/DNA polymerase	F_5	1500 (500 nM)
Multisite phosphorylation of Sic1		
Free Sic1 with i phosphorylations, $i = 0,\ldots,9$	Sic_i	Sic_0: 800 (267 nM)
S-Cdk bound Sic1 with i phos., $i = 0,\ldots,9$	Z_i	0
G1-Cdk bound Sic1–S-Cdk with i phos., $i = 0,\ldots,8$	Za_i	0
Cdc14 bound Sic1–S-Cdk with i phos., $i = 1,\ldots,9$	Zb_i	0
Cdks, Ddk and phosphatase		
Cln1,2–Cdc28 (G1-Cdk)	Cln	1500 (500 nM)
Clb5,6–Cdc28 (S-Cdk)	Cdk	600 (200 nM)
Dbf4–Cdc7 (Ddk)	Ddk	1600 (533 nM)
Phosphatase (Cdc14)	$Pase$	360 (120 nM)
Multisite phosphorylation of Sld2 and Sld3		
Free Sld2 with i phosphorylations, $i = 0,\ldots,7$	X_i	X_0: 660 (220 nM)
S-Cdk bound Sld2 with i phosphorylations, $i = 0,\ldots,6$	Xa_i	0
Cdc14 bound Sld2 with i phosphorylations, $i = 1,\ldots,7$	Xb_i	0
Free Sld3 with i phosphorylations, $i = 0,1,2$	Y_i	Y_0: 130 (43 nM)
S-Cdk bound Sld3 with i phosphorylations, $i = 0,1$	Ya_i	0
Cdc14 bound Sld3 with i phosphorylations, $i = 1,2$	Yb_i	0
Formation of 11-3-2 complex		
Dpb11	C_0	540 (180 nM)
Dpb11-Sld2p	C_1	0
Sld3p-Dpb11	C_2	0
Sld3p-Dpb11-Sld2p	C_3	0

Most values for the initial molecule numbers are taken from experimental data [27]. Although these numbers were determined in unsynchronized cell populations, they nevertheless provide a reasonable approximation for the molecule numbers in a single cell and a realistic estimation for the relative amounts of the different molecules.

The number of Cdc6 molecules was not determined experimentally and is assumed to be present in the nucleus in a medium concentration of $\sim$700 molecules per cell. The number of S-Cdk (Clb5,6-Cdc28) molecules is limited by the amount of its regulatory cyclins, Clb5 and Clb6, whose number was only measured for Clb5 protein (520 molecules per cell) [27]. The concentration of Clb6 was reported to be approximately ten times lower than the concentration of Clb5 [16], resulting in the estimation of $\sim$600 S-Cdk molecules per cell in total. The number of Cdc28 molecules, the catalytic subunit of S-Cdk and G1-Cdk, is not limiting, as approximately 6700 molecules are determined to be present in budding yeast cells [27]. The number of phosphatase molecules was reported to be higher than 8000 for Cdc14 [27]. A much lower number of 360 phosphatase molecules per nucleus is assumed in the mathematical model, since Cdc14 is known to be sequestered in the nucleolus during most of the cell cycle and to bind to a multitude of other substrates that are not part of this model [7, 8].

The number of early origins in the yeast cell is not precisely known and possibly varies among cells and cell cycles. In the mathematical model, a number of 190 early replication origins is assumed, which was determined by Lengronne et al. (2001) [46]. Moreover, the formation of two replication complexes per origin is required for the bidirectional DNA synthesis, resulting in an effective initial number of 380 origins of replication.

The initial number of molecules is converted into units of molecule concentrations in the mathematical model as follows:

$$[\text{molecule}] = \frac{\#\,\text{molecule}}{V_{\text{nucleus}} \cdot N_{\text{A}}}\ ,$$

with the volume of the nucleus, $V_{\text{nucleus}} = 5 \times 10^{-15}$ liter [38], and the Avogadro constant (number of molecules per mol), $N_{\text{A}} = 6.022 \times 10^{23}\ \text{mol}^{-1}$.

2.2.2 Mathematical equations

In the following, the balance equations for all intermediate molecule complexes and phosphorylation states are specified. The equations are separated into groups of reactions concerning the same molecule or describing a similar reaction type.

States of the replication origins

The assembly of the RC at the replication origin is described by a sequential transition of the origin through the states S_0 to S_9, which each represents a distinct combination of molecules bound to that origin (Figure 2.1 and Table 2.2).

Binding reactions are assumed to be generally reversible, and are specified by k_{on} and k_{off} rate constants (k_1 and k_{-1}, k_3 and k_{-3}, k_{20} and k_{-20}, k_{21} and k_{-21}, k_{22} and k_{-22}, k_{23} and k_{-23}, k_{24} and k_{-24}). An exception are the dissociation of Cdc6 and Cdt1 (rate constant k_9) and the dissociation

of the 11-3-2 activator complex (rate constant k_{25}), which follow the ATP-consuming, directed loading of the putative helicase, Mcm2-7, and the proposed catalytic replacement of the 11-3-2 complex in the course of the loading of the GINS–DNA polymerase complex, respectively.

The sequential phosphorylation of Mcm2-7 proteins in pre-RC by S-Cdk and Ddk (phosphorylation rate constants α_{20} and α_{21}) leads to the intermediate complexes between kinase and pre-RC, states S_{Cdk} and S_{Ddk}, and causes a directed, quasi-irreversible transition between the states S_3 and S_5.

When the RC is completed in state S_9, the fired origin is allowed to re-enter the DNA replication initiation network at state S_0 after a time delay of some minutes, which is specified by the rate constant k_{27}.

$$
\begin{aligned}
S_0\,'(t) &= +k_{27}S_9(t) - k_1S_0(t)F_1(t) + k_{-1}S_1(t) - k_{26}S_0(t)Cdk(t) + k_{-26}Sc_0(t)\\
S_1\,'(t) &= k_1S_0(t)F_1(t) - k_{-1}S_1(t) - k_3S_1(t)F_2(t) + k_{-3}S_2(t) - k_{26}S_1(t)Cdk(t) + k_{-26}Sc_1(t)\\
S_2\,'(t) &= k_3S_1(t)F_2(t) - k_{-3}S_2(t) - k_9S_2(t)\\
S_3\,'(t) &= k_9S_2(t) - k_{20}S_3(t)Cdk(t) + k_{-20}S_{Cdk}(t) - k_{26}S_3(t)Cdk(t) + k_{-26}Sc_3(t)\\
S_{Cdk}\,'(t) &= k_{20}S_3(t)Cdk(t) - k_{-20}S_{Cdk}(t) - \alpha_{20}S_{Cdk}(t)\\
S_4\,'(t) &= \alpha_{20}S_{Cdk}(t) - k_{21}S_4(t)Ddk(t) + k_{-21}S_{Ddk}(t) - k_{26}S_4(t)Cdk(t) + k_{-26}Sc_4(t)\\
S_{Ddk}\,'(t) &= k_{21}S_4(t)Ddk(t) - k_{-21}S_{Ddk}(t) - \alpha_{21}S_{Ddk}(t)\\
S_5\,'(t) &= \alpha_{21}S_{Ddk}(t) - k_{22}S_5(t)F_4(t) + k_{-22}S_6(t) - k_{26}S_5(t)Cdk(t) + k_{-26}Sc_5(t)\\
S_6\,'(t) &= k_{22}S_5(t)F_4(t) - k_{-22}S_6(t) - k_{23}S_6(t)C_3(t) + k_{-23}S_7(t) - k_{26}S_6(t)Cdk(t) + k_{-26}Sc_6(t)\\
S_7\,'(t) &= k_{23}S_6(t)C_3(t) - k_{-23}S_7(t) - k_{24}S_7(t)F_5(t) + k_{-24}S_8(t) - k_{26}S_7(t)Cdk(t) + k_{-26}Sc_7(t)\\
S_8\,'(t) &= k_{24}S_7(t)F_5(t) - k_{-24}S_8(t) - k_{25}S_8(t) - k_{26}S_8(t)Cdk(t) + k_{-26}Sc_8(t)\\
S_9\,'(t) &= k_{25}S_8(t) - k_{27}S_9(t) - k_{26}S_9(t)Cdk(t) + k_{-26}Sc_9(t)
\end{aligned}
\tag{2.1}
$$

S-Cdk phosphorylates ORC subunits, presumably Orc6 and Orc2 [59], with the rate constant α_{26}, contributing to prevent DNA rereplication. S-Cdk is assumed to bind to the origins in all states (rate constants k_{26} and k_{-26}), except in state S_2, in which S-Cdk is supposed to sterically interfere with Cdt1 [59, 14]. The kinetics of the transient complexes between S-Cdk and the origin are given by the balance equations, Sc_0, Sc_1 and Sc_3 to Sc_9:

$$
Sc_i\,'(t) = k_{26}S_i(t)Cdk(t) - k_{-26}Sc_i(t) - \alpha_{26}Sc_i(t), \quad i = 0,1,3,4,5,6,7,8,9
$$

The balance equations for the molecule complexes assembled at replication origins that are phosphorylated at ORC subunits, Sp_0, Sp_1 and Sp_3 to Sp_9, are:

$$
\begin{aligned}
Sp_0\,'(t) &= -k_1 Sp_0(t)F_1(t) + k_{-1}Sp_1(t) + \alpha_{26}Sc_0(t)\\
Sp_1\,'(t) &= k_1 Sp_0(t)F_1(t) - k_{-1}Sp_1(t) + \alpha_{26}Sc_1(t)\\
Sp_3\,'(t) &= \alpha_{26}Sc_3(t) - k_{20}Sp_3(t)Cdk(t) + k_{-20}Sp_{Cdk}(t)\\
Sp_{Cdk}\,'(t) &= k_{20}Sp_3(t)Cdk(t) - k_{-20}Sp_{Cdk}(t) - \alpha_{20}Sp_{Cdk}(t)\\
Sp_4\,'(t) &= \alpha_{26}Sc_4(t) + \alpha_{20}Sp_{Cdk}(t) - k_{21}Sp_4(t)Ddk(t) + k_{-21}Sp_{Ddk}(t)\\
Sp_{Ddk}\,'(t) &= k_{21}Sp_4(t)Ddk(t) - k_{-21}Sp_{Ddk}(t) - \alpha_{21}Sp_{Ddk}(t)\\
Sp_5\,'(t) &= \alpha_{26}Sc_5(t) + \alpha_{21}Sp_{Ddk}(t) - k_{22}Sp_5(t)F_4(t) + k_{-22}Sp_6(t)\\
Sp_6\,'(t) &= \alpha_{26}Sc_6(t) + k_{22}Sp_5(t)F_4(t) - k_{-22}Sp_6(t) - k_{23}Sp_6(t)C_3(t) + k_{-23}Sp_7(t)\\
Sp_7\,'(t) &= \alpha_{26}Sc_7(t) + k_{23}Sp_6(t)C_3(t) - k_{-23}Sp_7(t) - k_{24}Sp_7(t)F_5(t) + k_{-24}Sp_8(t)\\
Sp_8\,'(t) &= \alpha_{26}Sc_8(t) + k_{24}Sp_7(t)F_5(t) - k_{-24}Sp_8(t) - k_{25}Sp_8(t)\\
Sp_9\,'(t) &= \alpha_{26}Sc_9(t) + k_{25}Sp_8(t) - k_{27}Sp_9(t)
\end{aligned}
\tag{2.2}
$$

In the phosphorylated origin state Sp_i all reactions are assumed to occur with the same rate constant as in the unphosphorylated origin state S_i specified in equations (2.1). As already mentioned, the binding of the Cdt1–Mcm2-7 complex in state Sp_1 is impossible, since the state Sp_2 does not exist. The re-entering of completed, phosphorylated RCs, state Sp_9, to the initial state of the origins, Sp_0, is also not considered for RCs with phosphorylated ORC subunits, because these origins are inhibited to become licensed again. Due to the absence of state Sp_2, and the assumed irreversibility of the ORC phosphorylation by S-Cdk within the simulated time period, the formation of new, phosphorylated pre-RCs is impossible.

Molecules binding in licensing and firing phase

Equations (2.1) and (2.3) describe the binding of different molecules to the replication origins in distinct states: the licensing factors, Cdc6, Cdt1 and Mcm2-7 (F_1 to F_{2p}), in states S_0 and S_1, and the firing factors, Cdc45, GINS complex and DNA polymerase (F_4 and F_5), in states S_5 to S_7. Mcm2-7 molecules, Cdc45, GINS complex and the DNA polymerase (F_{2b}, F_4 and F_5) remain part of the RC, whereas the other molecules only associate transiently with the origins. The balance equations for the free concentrations of the binding factors, F_i, account for the binding, dissociation, degradation, and nuclear export of these molecules, but neglect protein synthesis as well as new import into the cell nucleus.

$$
\begin{aligned}
F_1\,'(t) &= -k_1F1(t)(S_0(t)+Sp_0(t))+k_{-1}(S_1(t)+Sp_1(t))+k_9S_2(t)-\delta_2F_1(t)f_{\mathrm{Cln}}(t)\\
F_{2a}\,'(t) &= -k_4F_{2a}(t)F_{2b}(t)+k_{-4}F_2(t)-k_5F_{2a}(t)F_{2bp}(t)+k_{-5}F_{2p}(t)+k_9S_2(t)-k_7F_{2a}(t)\\
F_{2b}\,'(t) &= -k_4F_{2a}(t)F_{2b}(t)+k_{-4}F_2(t)-\alpha_6F_{2b}(t)(f_{\mathrm{Cln}}(t)+\frac{Cdk(t)}{[\mathrm{Clb56}]})+k_{27}(S_9(t)+Sp_9(t))\\
F_{2bp}\,'(t) &= \alpha_6F_{2bp}(t)(f_{\mathrm{Cln}}(t)+\frac{Cdk(t)}{[\mathrm{Clb56}]})-k_5F_{2a}(t)F_{2bp}(t)+k_{-5}F_{2p}(t)-k_8F_{2bp}(t) \qquad (2.3)\\
F_2\,'(t) &= k_4F_{2a}(t)F_{2b}(t)-k_{-4}F_2(t)-k_3S_1(t)F_2(t)+k_{-3}S_2(t)-\alpha_6F_2(t)(f_{\mathrm{Cln}}(t)+\frac{Cdk(t)}{[\mathrm{Clb56}]})\\
F_{2p}\,'(t) &= \alpha_6F_2(t)(f_{\mathrm{Cln}}(t)+\frac{Cdk(t)}{[\mathrm{Clb56}]})+k_5F_{2a}(t)F_{2bp}(t)-k_{-5}F_{2p}(t)-k_8F_{2p}(t)\\
F_4\,'(t) &= -k_{22}F_4(t)(S_5(t)+Sp_5(t))+k_{-22}(S_6(t)+Sp_6(t))+k_{27}(S_9(t)+Sp_9(t))\\
F_5\,'(t) &= -k_{24}F_5(t)(S_7(t)+Sp_7(t))+k_{-24}(S_8(t)+Sp_8(t))+k_{27}(S_9(t)+Sp_9(t))
\end{aligned}
$$

The kinetic parameters for the binding reactions of the molecules to the replication origins in states S_i and Sp_i are specified already in equations (2.1) and (2.3). The G1-Cdk-dependent degradation of Cdc6 (F_1) is represented in the mathematical model by the rate constant, δ_2, multiplied by the time-dependent function for the G1-Cdk activation, $f_{\mathrm{Cln}}(t)$ (specified later in equation (2.16)). The formation of the complex consisting of Cdt1 (F_{2a}) and Mcm2-7 (F_{2b}) is described by the rate constants k_4 and k_{-4}. The phosphorylation of Mcm2-7 proteins, free or in complex with Cdt1, by G1-Cdk and S-Cdk is described by the phosphorylation rate constant, α_6, multiplied by the activation functions of G1-Cdk and S-Cdk, $(f_{\mathrm{Cln}}(t)+\frac{Cdk(t)}{[\mathrm{Clb56}]})$ (given in equations (2.16) and (2.18)). Phosphorylated Mcm2-7 proteins, Mcm2-7p, are assumed to have a lower binding affinity to Cdt1 (rate constants k_5 and k_{-5}). Free phosphorylated Mcm2-7p molecules (F_{2bp}) and phosphorylated Cdt1–Mcm2-7p complexes (F_{2p}) are exported from the nucleus with the same rate constant k_8. The nuclear export of free Cdt1 molecules is described by the rate constant k_7.

Multiple phosphorylation of Sld2 and Sld3

Experimental data by Tak et al. (2006) [77] indicated that threonine 84, which is essential for the binding of Sld2 to Dpb11, is phosphorylated with a time delay of some minutes compared to other serine and threonine phosphorylations. A conformational change in the Sld2 protein, occurring only after six random serine and threonine phosphorylations in Sld2 by S-Cdk, is supposed to be the reason for the delayed accessibility of threonine 84 for phosphorylation by S-Cdk [77, 80]. Simulation of the phosphorylation kinetics of six randomly phosphorylated serine and threonine residues leads to identical time courses for these phosphorylations, which has indeed been observed experimentally for two different residues (Figure 2.2(a)). Threonine 84 is assumed the last residue to be phosphorylated. This causes a delayed formation of fully phosphorylated Sld2 (Figure 2.2(a)).

For the binding of Sld3 to Dpb11, the phosphorylation of two residues in Sld3 by S-Cdk ap-

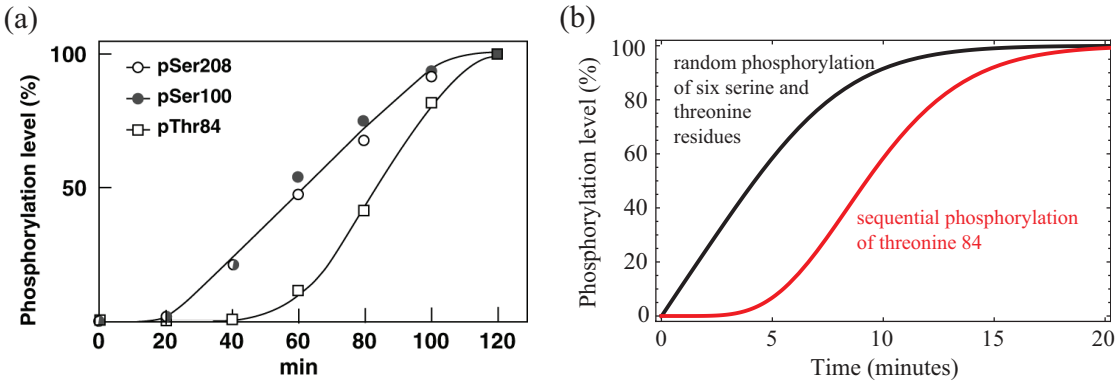


Figure 2.2: (a) The phosphorylation kinetics of three different residues in Sld2, measured *in vitro* by Tak et al. (2006) [77], indicate a simultaneous (random) phosphorylation of two serine residues and a delayed (sequential) phosphorylation of threonine 84 by S-Cdk. (b) The simulated phosphorylation kinetics of Sld2 can reproduce the experimentally observed, delayed phosphorylation of threonine 84 after six random phosphorylations. The time scale of phosphorylations in the simulation and *in vivo* is supposed to be faster than in the *in vitro* experiment.

pears to be crucial [81, 89], whose order is also assumed to be random.

In the mathematical model, the multiple random phosphorylation of Sld2 and Sld3 is simplified with the assumption that the binding and dissociation rate constants describing the complex formation of kinase and substrate, k_{14} and k_{-14}, as well as the phosphorylation rate constants, α_{14}, are equal for all randomly phosphorylated residues [70]. The same simplification is made for the multiple random dephosphorylation of Sld2 and Sld3 by the phosphatase, Cdc14, with the equivalent assumption, that the rate constants describing the complex formation of phosphatase and substrate, k_{15} and k_{-15}, as well as the dephosphorylation rate constants (β_{15}) are equal for all residues becoming dephosphorylated.

The high number of intermediate, partially phosphorylated states in a random phosphorylation scheme (Figure 2.3(a)) reduces to a sequential phosphorylation scheme (Figure 2.3(b)) by grouping all intermediate states with the same total number of phosphorylated residues together in one variable. In this case, the phosphorylation rate constants have to be slightly modified, by weighting them by the total number of phosphorylations that are possible in the next step, which is equal to the total number of up to then unphosphorylated residues in the molecule. In the case of a multiple dephosphorylation, the dephosphorylation rate constants have to be weighted by the number of next possible dephosphorylation steps, which is equal to the number of phosphorylated residues.

Using this simplification, the number of differential equations needed to describe the six random phosphorylations in Sld2 reduces from 64 to seven differential equations, X_0 to X_6. The last sequential phosphorylation of threonine 84 in Sld2, is described by an additional equation, X_7. This phosphorylation step is assumed to have the same binding and dissociation rate constants of kinase, S-Cdk (k_{16} and k_{-16}), and phosphatase, Cdc14 (k_{17} and k_{-17}), but the phosphorylation

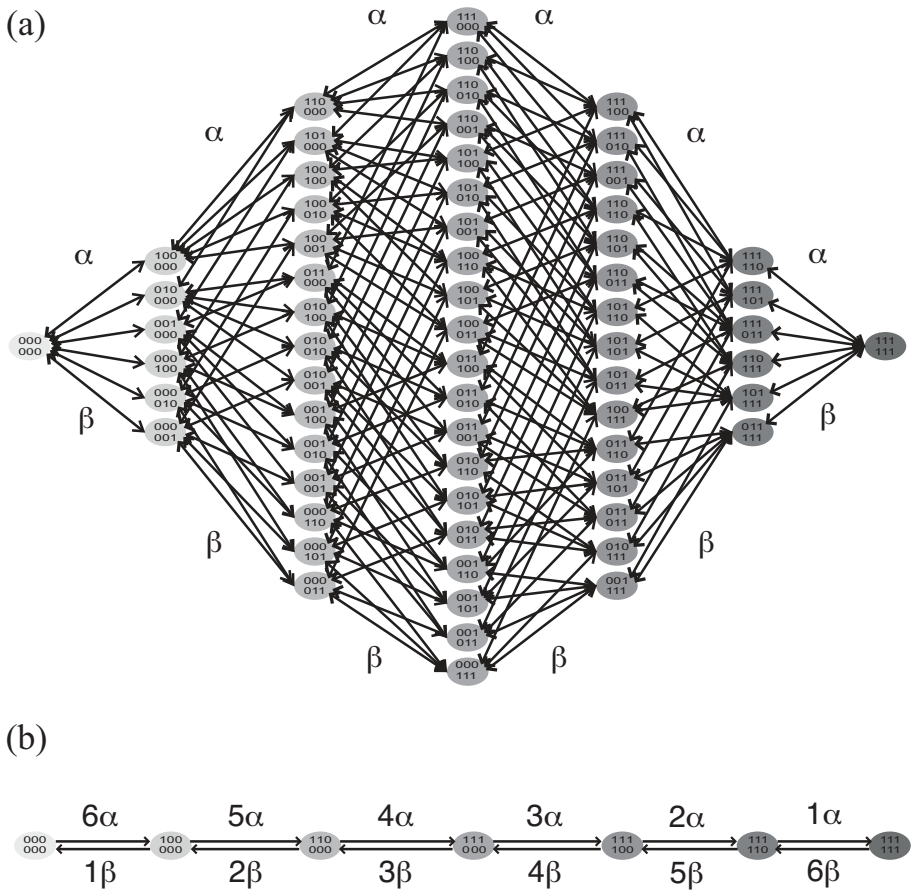


Figure 2.3: (a) A random (de)phosphorylation of six residues in a molecule comprises 64 different, partially phosphorylated states (α: phosphorylation rate constant, β: dephosphorylation rate constant). (b) This random (de)phosphorylation of six residues reduces to a sequential phosphorylation scheme with weighted (de)phosphorylation rate constants, α and β, if these are assumed to be equal for all (de)phosphorylations.

and dephosphorylation rate constants (α_{16} and β_{17}) are taken to be different.

$$
\begin{aligned}
X_0\,'(t) &= -k_{14}X_0(t)Cdk(t) + k_{-14}Xa_0(t) + \beta_{15}Xb_1(t) \\
X_i\,'(t) &= (6-(i-1))\alpha_{14}Xa_0(t) - k_{14}X_1(t)Cdk(t) + k_{-14}Xa_1(t) + (i+1)\beta_{15}Xb_2(t) \\
&\quad - k_{15}X_1(t)Pase(t) + k_{-15}Xb_1(t), \qquad i = 1,\ldots,5 \\
X_6\,'(t) &= \alpha_{14}Xa_5(t) - k_{16}X_6(t)Cdk(t) + k_{-16}Xa_6(t) + \beta_{17}Xb_7(t) - k_{15}X_6(t)Pase(t) + k_{-15}Xb_6(t) \\
X_7\,'(t) &= \alpha_{16}Xa_6(t) - k_{17}X_7(t)Pase(t) + k_{-17}Xb_7(t) - k_{18}X_7(t)(C_0(t)+C_2(t)) \\
&\quad + k_{-18}(C_1(t)+C_3(t))
\end{aligned}
\tag{2.4}
$$

To account for the kinetics of the intermediate complexes between kinase, S-Cdk, and substrate, Sld2, and between phosphatase, Cdc14, and substrate, Sld2, seven additional balance equations are needed for each complex, Xa_0 to Xa_6, and Xb_1 to Xb_7.

$$\begin{aligned} Xa_i\,'(t) &= k_{14}X_i(t)Cdk(t) - k_{-14}Xa_i(t) - (6-i)\alpha_{14}Xa_i(t), \quad i = 0,\ldots,5 \qquad (2.5)\\ Xa_6\,'(t) &= k_{16}X_6(t)Cdk(t) - k_{-16}Xa_6(t) - \alpha_{16}Xa_6(t)\\ Xb_i\,'(t) &= k_{15}X_i(t)Pase(t) - k_{-15}Xb_i(t) - i\beta_{15}Xb_i(t), \quad i = 1,\ldots,6 \qquad (2.6)\\ Xb_7\,'(t) &= k_{17}X_7(t)Pase(t) - k_{-17}Xb_7(t) - \beta_{17}Xb_7(t) \end{aligned}$$

The number of differential equations for the multiple phosphorylation of Sld3 by S-Cdk reduces accordingly from four to three equations, Y_0 to Y_2.

$$\begin{aligned} Y_0\,'(t) &= -k_{14}Y_0(t)Cdk(t) + k_{-14}Ya_0(t) + \beta_{15}Yb_1(t)\\ Y_1\,'(t) &= 2\alpha_{14}Ya_0(t) - k_{14}Y_1(t)Cdk(t) + k_{-14}Ya_1(t) + 2\beta_{15}Yb_2(t) - k_{15}Y_1(t)Pase(t)\\ &\quad + k_{-15}Yb_1(t) \qquad (2.7)\\ Y_2\,'(t) &= \alpha_{14}Ya_1(t) - k_{15}Y_2(t)Pase(t) + k_{-15}Yb_2(t) - k_{19}Y_2(t)(C_0(t)+C_1(t))\\ &\quad + k_{-19}(C_2(t)+C_3(t)) \end{aligned}$$

Two additional balance equations each describe the kinetics of the intermediate complexes between S-Cdk and Sld3, Ya_0, Ya_1, and between Cdc14 and Sld3, Yb_1, Yb_2.

$$Ya_i\,'(t) = k_{14}Y_i(t)Cdk(t) - k_{-14}Ya_i(t) - (i+1)\alpha_{14}Ya_i(t), \quad i = 0,1 \qquad (2.8)$$

$$Yb_i\,'(t) = k_{15}Y_i(t)Pase(t) - k_{-15}Yb_i(t) - \beta_{15}Yb_i(t), \quad i = 1,2 \qquad (2.9)$$

The binding, dissociation, and (de)phosphorylation rate constants for the reactions with Sld3 are taken to be the same as for the multiple random phosphorylation of Sld2.

11-3-2 complex formation

Completely phosphorylated Sld2 and Sld3 bind to Dpb11 and form the 11-3-2 complex (C_3). The precise order of this complex assembly is still unclear [80]. We assume a random order with two intermediate complexes, Sld2–Dpb11 (C_1) and Sld3–Dpb11 (C_2). The binding and dissociation rate constants for Sld2 to Dpb11 (k_{18} and k_{-18}) and Sld3 to Dpb11 (k_{19} and k_{-19}) are each assumed to be independent of the binding order, as a first or second molecule, to Dbp11. The kinetic parameters for the binding and dissociation of the 11-3-2 activator complex to the origins (k_{23} and k_{-23} and k_{25}) are already specified in the balance equations for the states of the origins (equations (2.1) and (2.3)).

$$
\begin{aligned}
C_0\,'(t) &= -C_0(t)(k_{18}X_7(t)+k_{19}Y_2(t))+k_{-18}C_1(t)+k_{-19}C_2(t)\\
C_1\,'(t) &= k_{18}X_7(t)C_0(t)-k_{19}Y_2(t)C_1(t)-k_{-18}C_1(t)+k_{-19}C_3(t)\\
C_2\,'(t) &= k_{19}Y_2(t)C_0(t)-k_{18}X_7(t)C_2(t)-k_{-19}C_2(t)+k_{-18}C_3(t)\\
C_3\,'(t) &= k_{19}Y_2(t)C_1(t)+k_{18}X_7(t)C_2(t)-k_{-19}C_3(t)-k_{-18}C_3(t)-k_{23}C_3(t)(S_6(t)+Sp_6(t))\\
&\quad +k_{-23}(S_7(t)+Sp_7(t))+k_{25}(S_8(t)+Sp_8(t))
\end{aligned}
\tag{2.10}
$$

Multiple phosphorylation and degradation of the S-Cdk inhibitor, Sic1

The stoichiometric inhibitor of S-Cdk, Sic1, is phosphorylated randomly by G1-Cdk at multiple residues before it is recognized by the SCF^{Cdc4} complex and degraded via ubiquitin-dependent proteolysis. Subsequently, active S-Cdk becomes free [58, 20]. Sic1 is assumed to bind to S-Cdk with a very high affinity (rate constants k_{12} and k_{-12}), which is independent of the phosphorylation state of Sic1. Nine phosphorylation sites are considered in Sic1 in total, whereas already after any combination of six up to nine Sic1 phosphorylations by G1-Cdk [58, 20], degradation of Sic1 (rate constant δ_{13}) is rendered possible in the mathematical model.

$$
\begin{aligned}
Sic_i\,'(t) &= -k_{12}Cdk(t)Sic_i(t)+k_{-12}Z_i(t), \quad i=0,\ldots,5\\
Sic_i\,'(t) &= -k_{12}Cdk(t)Sic_i(t)+k_{-12}Z_i(t)-\delta_{13}Sic_i(t), \quad i=6,\ldots,9
\end{aligned}
\tag{2.11}
$$

The mathematical description of the multiple random phosphorylation of the Sic1–S-Cdk complex is simplified in the model as in the case of Sld2 and Sld3. The rate constants for the kinase, G1-Cdk, to substrate, Sic1–S-Cdk, binding and dissociation, k_{10} and k_{-10}, and the phosphorylation rate constants, α_{10}, as well as the rate constants for the phosphatase, Cdc14, to substrate, Sic1–S-Cdk, binding and dissociation, k_{11} and k_{-11}, and the dephosphorylation rate constants, β_{11}, are taken to be the same for all residues, that become (de)phosphorylated. In the case of the nine random phosphorylations in Sic1, the number of phosphorylation states that need to be considered in the mathematical model reduces from $2^9 = 512$ to only 10, Z_0 to Z_9.

$$
\begin{aligned}
Z_0\,'(t) &= k_{12}Cdk(t)Sic_0(t)-k_{-12}Z_0(t)-k_{10}Cln(t)f_{\mathrm{Cln}}(t)Z_0(t)+k_{-10}Za_0(t)+\beta_{11}Zb_1(t)\\
Z_i\,'(t) &= k_{12}Cdk(t)Sic_i(t)-k_{-12}Z_i(t)+(9-(i-1))\alpha_{10}Za_{i-1}(t)-k_{11}Z_i(t)Pase(t)\\
&\quad +k_{-11}Zb_i(t)-k_{10}Cln(t)f_{\mathrm{Cln}}(t)Z_i(t)+k_{-10}Za_i(t)+(i+1)\beta_{11}Zb_{i+1}(t), \quad i=1,\ldots,8\\
Z_9\,'(t) &= k_{12}Cdk(t)Sic_9(t)-k_{-12}Z_9(t)+\alpha_{10}Za_8(t)-k_{11}Z_9(t)Pase(t)+k_{-11}Zb_9(t)
\end{aligned}
\tag{2.12}
$$

Nine differential equations are each needed to describe the intermediate complexes between G1-Cdk and Sic1–S-Cdk, Za_0 to Za_8, and between the phosphatase, Cdc14, and Sic1–S-Cdk, Zb_1 to Zb_9.

$$Za_i{}'(t) = k_{10}Cln(t)f_{\mathrm{Cln}}(t)Z_i(t) - k_{-10}Za_i(t) - (9-i)\alpha_{10}Za_i(t), \quad i = 0,\ldots,8 \tag{2.13}$$
$$Zb_i{}'(t) = k_{11}Z_i(t)Pase(t) - k_{-11}Zb_i(t) - i\beta_{11}Zb_i(t), \quad i = 1,\ldots,9$$

Cell cycle-dependent kinases, G1-Cdk and S-Cdk, phosphatase, and Ddk

G1-Cdk becomes active in late G1 phase and is responsible for the termination of the licensing phase and the activation of S-Cdk. In the mathematical model, a sigmoid time-dependent function describes the switch-like activation of G1-Cdk, which is due to the experimentally observed positive feedback loop in the expression of the G1 phase cyclins, Cln1 and Cln2 [74], the regulatory subunits of G1-Cdk. More precisely, G1-Cdk is assumed to become active approximately 20 minutes after the start of origin licensing (time shift $t_0 = 1500$ seconds) and to reach its maximal activity within around ten minutes (time constant $\lambda = 0.0075$ 1/second) in the mathematical model.
Experimental data of Cln2 synthesis indicate, that G1-Cdk becomes active 20 to 60 minutes after birth in daughter cells [74].

$$f_{\mathrm{Cln}}(t) = \frac{1}{2}(1 + \tanh(\lambda(t - t_0))) \tag{2.14}$$

$$= \frac{1}{2}(1 + \tanh(0.0075(t - 1500)))$$

The balance equation for the free G1-Cdk concentration, $Cln(t)$, considers the complex formation with Sic1–S-Cdk and its phosphorylation reaction (equations (2.13) and (2.14)). In the ODE system, all kinetic rate constants of reactions involving free G1-Cdk are modified by the activation function of G1-Cdk, $f_{\mathrm{Cln}}(t)$ (equation (2.16)).

$$Cln'(t) = -k_{10}Cln(t)f_{\mathrm{Cln}}(t)\sum_{i=0}^{8} Z_i(t) + k_{-10}\sum_{i=0}^{8} Za_i(t) + \alpha_{20}(\sum_{i=0}^{8}(9-i)Za_i(t)) \tag{2.15}$$

$$\tag{2.16}$$

The balance equation for the free S-Cdk concentration, $Cdk(t)$, considers the binding reactions of S-Cdk to Mcm2-7 proteins, to ORC subunits in the different origin states, and to Sic1, Sld2, and Sld3 in all intermediate phosphorylation states and with the corresponding phosphorylation reactions. The kinetic parameters of these reactions have already been specified before.

$$
\begin{aligned}
Cdk'(t) = &-k_{12}Cdk(t)\sum_{i=0}^{9}Sic_i(t) + k_{-12}\sum_{i=0}^{9}Z_i(t) \\
&- k_{14}Cdk(t)(\sum_{i=0}^{6}X_i(t) + \sum_{i=0}^{1}Y_i(t)) + k_{-14}(\sum_{i=0}^{6}Xa_i(t) + \sum_{i=0}^{1}Ya_i(t)) \\
&+ \alpha_{14}(\sum_{i=0}^{5}(6-i)Xa_i(t) + \sum_{i=0}^{1}(2-i)Ya_i(t)) + \alpha_{16}Xa_6(t) \\
&- k_{26}Cdk(t)(\sum_{i=0}^{1}S_i(t) + \sum_{i=3}^{9}S_i(t)) + k_{-26}(\sum_{i=0}^{1}Sc_i(t) + \sum_{i=3}^{9}Sc_i(t)) + \alpha_{26}(\sum_{i=0}^{1}Sc_i(t) + \sum_{i=3}^{9}Sc_i(t)) \\
&- k_{20}Cdk(t)(S_3(t) + Sp_3(t)) + k_{-20}(S_{Cdk}(t) + Sp_{Cdk}(t)) + \alpha_{20}(S_{Cdk}(t) + Sp_{Cdk}(t))
\end{aligned}
\tag{2.17}
$$

The balance equation for the free concentration of the phosphatase, $Pase(t)$, accounts for the binding reactions of the phosphatase to Sic1, Sld2 and Sld3 in all intermediate phosphorylation states and with the corresponding dephosphorylation reactions. The kinetic parameters of these reactions are specified in equations (2.5) to (2.10) and equations (2.13) to (2.15).

$$
\begin{aligned}
Pase'(t) = &-k_{11}Pase(t)\sum_{i=1}^{9}Z_i(t) + k_{-11}\sum_{i=1}^{9}Zb_i(t) + \beta_{11}(\sum_{i=1}^{9}iZb_i(t)) \\
&- k_{15}Pase(t)(\sum_{i=1}^{7}X_i(t) + \sum_{i=1}^{2}Y_i(t)) + k_{-15}(\sum_{i=1}^{7}Xb_i(t) + \sum_{i=1}^{2}Yb_i(t)) \\
&+ \beta_{15}(\sum_{i=1}^{6}iXb_i(t) + \sum_{i=1}^{2}iYb_i(t)) + \beta_{17}Xb_7(t)
\end{aligned}
\tag{2.18}
$$

The balance equation for the free Ddk concentration, $Ddk(t)$, considers the binding of Ddk to Mcm2-7 proteins and the phosphorylation of the Mcm2-7 molecules. The kinetic parameters of these reactions are specified in equations (2.1) and (2.3).

$$
\begin{aligned}
Ddk'(t) = &-k_{21}Ddk(t)(S_4(t) + Sp_4(t)) + k_{-21}(S_{Ddk}(t) + Sp_{Ddk}(t)) \\
&+ \alpha_{21}(S_{Ddk}(t) + Sp_{Ddk}(t))
\end{aligned}
\tag{2.19}
$$

2.3 Functional parameterizations of the mathematical model

To parametrize the mathematical model, parameter sets that have the ability to reproduce the functionality of the biological system are generated systematically. Firstly, functional properties characterizing the biological system are formulated mathematically allowing to quantify and compare the systems functionality of different model parameterizations. Secondly, biochemically realistic parameter ranges are specified for the various types of kinetic rate constants.

Finally, random parameter sets are generated within the biochemically allowed ranges and subsequently optimized, in order to maximize the systems functionality and to account for the observed degree of functionality.

2.3.1 Definition of functional systems properties

The kinetics of origin firing has to fulfill certain conditions in order to be physiologically functional. In the mathematical model, the kinetics of origin firing is described by the firing rate, $f(t)$, which gives the number of origins activated per time interval:

$$f(t) = k_{25}(S8(t) + S8p(t)). \tag{2.20}$$

$f(t)$ is the principle output of simulating the dynamics of the replication initiation network and can be compared directly to experimental measurements of activation times of replication origins in the budding yeast genome (see next chapter).

The mathematical formulation of the kinetics of origin firing allows its quantification by means of four functional systems properties that are defined through the zeroth, first and second momentum of the distribution of origin firing times, $f(t)$, (Figure 2.4):

- The *number of firing origins*,

$$N = \int_0^\infty f(t) \mid_{k_{27}=0} \mathrm{d}t,$$

  comprises all origins that are activated in one round of the cell cycle. Reactivated origins that have already been fired once and thus initiate DNA rereplication are not considered. A high number of activated origins is supposed to be advantageous, since then the multitude of origins is exploited most efficiently and the duplication of the genome becomes faster.

- The *number of rereplicating origins*,

$$\rho = \int_0^\infty f(t)\,\mathrm{d}t - N,$$

  is determined as the difference between the total number of origin firing events and the number of origins that are fired for the first time, N, in one cell cycle. As the number of origins is converted into units of concentration in the mathematical model and thus is a continuous variable, the number of rereplicating origins, ρ, can be determined even if it is below one. In this case, ρ has the interpretation of a probability that one origin out of 190 origins in total is activated a second time in one cell cycle and causes DNA rereplication.

- The *mean time to origin firing*,

$$\tau = \frac{\int_0^\infty t\, f(t)\,\mathrm{d}t}{\int_0^\infty f(t)\,\mathrm{d}t},$$

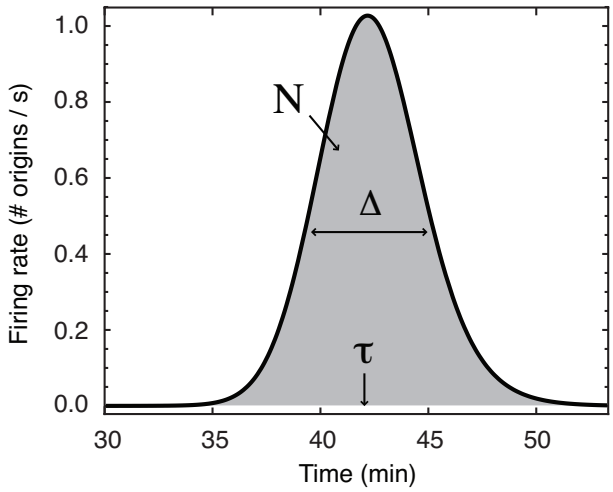


Figure 2.4: The firing rate, $f(t)$ is defined as the number of origins activated per time interval. It is characterized by the number of fired origins, N, the mean firing time, τ, and the standard deviation of firing, Δ, providing a quantitative description of the kinetics of origin firing.

is defined as the center of mass of the firing rate, $f(t)$, and gives the average duration of the RC formation at the early origins, counted from the beginning of origin licensing to the time of activation of replication origins.

- The *duration of origin firing*,

$$\Delta = 2\sqrt{\frac{\int_0^\infty t^2 f(t)\,\mathrm{d}t}{\int_0^\infty f(t)\,\mathrm{d}t} - \tau^2},$$

is defined as twice the standard deviation of the firing rate, $f(t)$, and characterizes the temporal distribution of the origin firing events. A short duration of origin firing, Δ, is supposed to efficiently exploit the multitude of replication origins enabling a preferably parallel and fast replication of the genome.

A parametrized mathematical model for the initiation of DNA replication is considered to reproduce the physiological required functionality, if the following conditions in terms of functional system properties are fulfilled:

- the number of firing origins, N, is higher than 180.5 origins, such that at least a fraction of 95% of the initial origins is activated,
- the number of rereplicating origins, ρ, is smaller than 0.01 origins, such that an origin is activated twice in only every 100^{th} cell cycle on average,

- the mean time to firing, τ, is below 85 minutes, and
- the duration of firing, Δ, is shorter than 25 minutes.

2.3.2 Biochemically allowed parameter ranges

The kinetic parameters describing the different types of reactions, that are binding, dissociation, phosphorylation and dephosphorylation reactions, are biochemically restricted. In the mathematical model, the values of the kinetic parameters are constrained within the following ranges.

Binding rate constant:	k_x	$\in$	$\left[10^3,\ldots,5\times10^5\right]$ $\mathrm{M}^{-1}\,\mathrm{s}^{-1}$
Dissociation rate constant:	k_{-x}	$\in$	$\left[k_x\times10^{-8},\ldots,k_x\times10^{-6}\right]$ s^{-1}
Phosphorylation rate constant:	α_x	$\in$	$[0.01,\ldots,1]$ s^{-1}
Dephosphorylation rate constant:	β_x	$\in$	$[0.05,\ldots,5]$ s^{-1}

Binding rate constants are assumed to be not faster than the diffusion limit for bimolecular reactions [61]. The dissociation constants, $K_\mathrm{D} = k_{-x}/k_x$, are considered to not be smaller than 10^{-8} M, which is reported to be an appropriate bound for the affinity of protein-protein or protein-DNA interactions [26]. Phosphorylation rate constants are assumed to be in the second to sub-second range and dephosphorylation reactions are taken to be slightly faster than phosphorylation reactions [43].

As the reactions have to be rather fast to account for the observed kinetics of origin firing, the kinetic rate constants are considered to be minimally 0.01 to 0.001 times slower than their maximum biochemically allowed rate, limiting the possible ranges for the kinetic parameter values to 2 to 3 orders of magnitude.

2.3.3 Determination of functional parameter sets through optimization

Functional parameter sets are generated by, firstly, assigning the kinetic parameter values of the licensing and S-Cdk activation modules according to experimental, quantitative and qualitative, observations. In particular, we considered that licensing of early origins was completed before S-Cdk became active. The dissociation constant for the binding of Cdc6 to the origins [25] and the Cdc6 degradation rate constant [23] have been measured, and values for the binding rate constants of Sic1 and S-Cdk were reported [5, 6]. All other rate constants are chosen within the biochemically allowed ranges. Secondly, several parameter sets are generated with fixed, identical kinetic parameters for the licensing and S-Cdk activation modules (Table 2.3), and different, randomly drawn kinetic parameters for the firing and 11-3-2 activator modules (see Figure 2.1 for the model submodules). The values of all kinetic parameters are constraint within biochemically given ranges and are assigned randomly on a logarithmic scale to ensure an equal distribution of the parameter values, across the orders of magnitudes, in the different parameter sets.

Table 2.3: Fixed values of the kinetic rate constants in the licensing and S-Cdk activation module in all parameter sets.

Parameter	Symbol	Value	Unit	Reference
Licensing module				
Binding of Cdc6	k_1	2×10^5	$(\mathrm{Ms})^{-1}$	[25]
Dissociation of Cdc6	k_{-1}	0.2	s^{-1}	[25]
Degradation of Cdc6	δ_2	0.005	s^{-1}	[23]
Binding of Cdt1-Mcm2-7	k_3	1×10^5	$(\mathrm{Ms})^{-1}$	
Dissociation of Cdt1-Mcm2-7	k_{-3}	0.05	s^{-1}	
Binding of Cdt1 and Mcm2-7	k_4	2×10^5	$(\mathrm{Ms})^{-1}$	
Dissociation of Cdt1 and Mcm2-7	k_{-4}	0.02	s^{-1}	
Binding of Cdt1 and Mcm2-7p	k_5	2×10^3	$(\mathrm{Ms})^{-1}$	
Dissociation of Cdt1 and Mcm2-7p	k_{-5}	2	s^{-1}	
Phosphorylation of (Cdt1-)Mcm2-7	α_6	0.002	s^{-1}	
Export of Cdt1	k_7	0.005	s^{-1}	
Export of (Cdt1-)Mcm2-7p	k_8	0.005	s^{-1}	
Dissociation of Cdc6, Cdt1	k_9	0.167	s^{-1}	
S-Cdk activation module				
Binding of G_1-Cdk and S-Cdk-Sic1	k_{10}	5×10^5	$(\mathrm{Ms})^{-1}$	
Dissociation of G_1-Cdk and S-Cdk-Sic1	k_{-10}	5	s^{-1}	
Phosphorylation of S-Cdk-Sic1	α_{10}	0.1	s^{-1}	
Binding of Cdc14 and S-Cdk-Sic1	k_{11}	5×10^4	$(\mathrm{Ms})^{-1}$	
Dissociation of Cdc14 and S-Cdk-Sic1	k_{-11}	0.5	s^{-1}	
Dephosphorylation of S-Cdk-Sic1	β_{11}	2.5	s^{-1}	
Binding of Sic1 and S-Cdk	k_{12}	2.5×10^6	$(\mathrm{Ms})^{-1}$	[6]
Dissociation of Sic1 and S-Cdk	k_{-12}	9.2×10^{-3}	s^{-1}	[5]
Degradation of Sic1(ip), $i = 6, \ldots, 9$	δ_{13}	1	s^{-1}	

Simulations with the mathematical model revealed that parameter sets with randomly generated parameter values, although constraint within biochemically reasonable ranges, are generally not able to reproduce a physiologically functional kinetic behavior. In most cases, the timing of origin firing was by far too slow resulting in an average mean time to origin firing, τ, and a duration of firing, Δ, of longer than 15 hours. In only 9 out of 10000 randomly generated parameter sets, the kinetics of origin firing was found to be functional, as defined above. However, the average functional systems properties of these parameter sets (activated origins $N = 185.6$ origins, rereplicating origins $\rho = 3.9 \times 10^{-5}$ origins, mean time to origin firing $\tau = 63.9$ minutes, and duration of firing $\Delta = 19.6$ minutes) reveal that the kinetics of origin firing is still not fast enough to account for the experimentally observed distribution of origin firing times (Figure 3.1 (b) and (c)).

This indicates a need for a more systematic determination of functional parameter sets, which

is realized by an optimization of the systems functionality as a function of the kinetic parameters.

Optimization of the systems functionality

All biological systems are subject to natural evolution, in which mutations and selection lead to an optimization of the functional behavior of the system. Hence, the computational optimization of the functionality of a biological system, in order to determine its kinetic parameters, is an appropriate approach, which has previously been applied in the analysis of metabolic networks, e.g. to determine optimal enzyme concentrations or maximal enzyme activities [34].

The mathematical model for the initiation of DNA replication in budding yeast revealed that the kinetics of origin firing generated with random parameter sets was much too slow compared to the experimental observations. To enhance the functionality of the system, that is to accelerate the kinetics and the coherence of origin firing, the duration of origin firing, Δ, is chosen as an objective function in the optimization process, in which it is minimized as a function of the kinetic parameters. As the duration of firing, Δ, and the mean time to firing, τ, are interdependent functional systems properties (shown later), the minimization of the duration of firing, Δ, implies an optimization of the overall timing of origin firing, characterized by Δ and τ. The timing of origin firing is supposed to be independent of reactions in the licensing and S-Cdk activation modules [22, 10]. Hence, only the 23 kinetic parameters of the firing and 11-3-2 activator module (out of 51 kinetic parameters in total) are subject to an optimization of the duration of origin firing, Δ, for a given S-Cdk activation kinetics:

$$\min \Delta(\vec{p}) = \min \Delta(\{p_1, p_2, \ldots, p_{23}\}).$$

An evolutionary optimization algorithm, which relies on a stochastic optimization approach, is chosen, because these algorithms have demonstrated to work well for complex optimization problems [4, 57]. In contrast to deterministic optimization methods, like gradient-based methods that use first or second order derivatives and are only good in finding local optima, stochastic optimization methods are applied to search for global optima in systems with multiple local minima or maxima. Moreover, evolutionary optimization algorithms are easy to implement, robust and inherently parallel. Contrariwise, they are computationally rather expensive.
The minimization of the duration of firing, Δ, was implemented in `Mathematica 7.0` [86] using the function, `NMinimize`, with the specific method `DifferentialEvolution`. The biochemically allowed ranges for the kinetic parameter values as well as additional requirements to the systems functionality were integrated as constraints in the algorithm.

Evolutionary optimization algorithm

1. The algorithm starts with a population of m randomly generated parameter sets,

$$\{\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_j, \ldots, \vec{p}_m\},$$

where typically $m >> n$, with n being the number of parameter values to be optimized (here: $n = 23$ and $m = 1000$, see Table 2.4).

2. In each iteration, the algorithm generates a new population of m parameter sets maintaining the total number of m parameter sets.
   The j^{th} new parameter set, $\vec{p}_{j_{new}}$, is created as follows:

   2.a) Three old parameter sets, $\vec{p}_u$, $\vec{p}_v$ and $\vec{p}_w$, are randomly selected from the population of m parameter sets.

   2.b) An auxiliary parameter set, $\vec{p}_s$, is formed,

   $$\vec{p}_s = \vec{p}_w + s\,(\vec{p}_u - \vec{p}_v),$$

   where s is a real scaling factor (here: $s = 0.6$, see Table 2.4).

   2.c) The j^{th} new parameter set, $\vec{p}_{j_{new}}$, is then composed of parameter values, $p_{j_{new},i}$, from the old, $\vec{p}_{j_{old}}$, and the auxiliary, $\vec{p}_s$, parameter set:

   $$p_{j_{new},i} = \begin{cases} p_{s,i} & \text{with probability } \eta \\ p_{j_{old},i} & \text{with probability } 1-\eta \end{cases} \quad \text{, for } i = 1,\ldots,23,$$

   where η is a real number between 0 and 1 (here: $\eta = 0.5$, see Table 2.4).

   2.d) The objective function, Δ, is evaluated for both parameter sets, $\vec{p}_{j_{new}}$ and $\vec{p}_{j_{old}}$, and the old parameter set, $\vec{p}_{j_{old}}$, is replaced by the new parameter set, $\vec{p}_{j_{new}}$, in the population, if:

   $$\Delta(\vec{p}_{j_{new}}) < \Delta(\vec{p}_{j_{old}}).$$

The options of the `DifferentialEvolution` method and the appropriate values chosen here are listed in Table 2.4.

Fully optimized parameter sets

Ten parameter sets were generated by minimizing the duration of origin firing, Δ, without limiting the total number of iteration steps in the algorithm. The values of the kinetic rate constants and the resulting functional systems properties differ in these fully optimized parameter sets, indicating that the algorithm has stopped automatically before it has converged to a global optimal solution. The generation of a fully optimized parameter set took on average 5 days on a 2.4 GHz Intel Mac. The shortening of the duration of origin firing during the iterations indicates that, usually, a reasonably short duration of origin firing is found already after a few iteration steps. Thus, to reduce the time spent on the generation of a functional parameter set, the total number of iteration steps was limited in the evolutionary optimization algorithm. This procedure is justifiable by the fact that the average functional systems properties obtained with the fully optimized parameter sets (activated origins $N = 181.2 \pm 1.6$ origins, rereplicating origins $\rho = 0.004 \pm 0.003$ origins, mean time to origin firing $\tau = 42.8 \pm 1.9$ minutes and duration of firing $\Delta = 5.9 \pm 1.5$ minutes) are generally not improved compared to the functional systems

Table 2.4: Options and values specified in the parameter optimization algorithm with the `DifferentialEvolution` method in `Mathematica 7.0` [86].

Option	Value	Explanation
`InitialPoints`	`Automatic`	Set of initial parameter sets
`SearchPoints` (m)	1000	Size of the population used for evolution
`RandomSeed`	$0,\ldots,1000$	Starting value for the random number generator
`ScalingFactor` (s)	0.6	Scale applied to the difference vector
`CrossProbability` (η)	0.5	Probability to take a parameter from $\vec{p}_s$
`Tolerance`	0.001	Tolerance for accepting constraint violations
`PenaltyFunction`	$d \cdot 10^{4 \cdot i}$	Function applied to penalize invalid parameter values outside constraints (d =distance from allowed value, i =number of iteration)
`PostProcess`	`True`	Whether to post-process using local search methods

properties obtained with sufficiently optimized parameter sets. The values of the functional systems properties generated by both types of kinetic parameter sets are in the same range (compare with Figure 3.2).

Sufficiently optimized parameter sets

For the systematic generation of functional parameter sets, the minimization of the duration of origin firing, Δ, was stopped after two iteration steps of the evolutionary optimization algorithm, which took on average 10 hours on a 2.4 GHz Intel Mac. Usually, a sufficiently fast parameter set was found already after two iteration steps. The algorithm was ran several times, each with a different random seed for the random number generator. In this way, a population of different, but equally functional parameter sets was obtained that enable to compare the functionality in the different parameter sets.

A total of 109 functional parameter sets were generated with fixed parameter values for the licensing and S-Cdk activation modules (Table 2.3) and optimized parameter values for the firing and 11-3-2 activator modules. The median, 1/6- and 5/6-quantile values of the obtained kinetic parameters in the different functional parameter sets are summarized in Figure 2.5.
The optimized dissociation and binding rate constants for the protein-protein and protein-DNA interactions attained preferred values in the admissible parameter sets, while the values of the phosphorylation and, in particular, the dephosphorylation rate constants seem to vary considerably in the different functional parameter sets. A hierarchy is identifiable in the binding affinities of S-Cdk to its substrates. The binding of S-Cdk to Orc subunits turned out to be typically less affine than the binding of S-Cdk to Mcm2-7 proteins, which, in turn, is still weaker than the affinity of S-Cdk to Sld2 and Sld3. In almost all functional parameter sets, the highest affinity is

detected for the binding of the 11-3-2 activator to the pre-ICs.

One parameter set is selected as a reference parameter set from all 109 functional parameter sets. In the following, it is used in simulations of characteristic perturbations of S-Cdk and of other experimental situations. The optimized values of the kinetic parameters in the firing and 11-3-2 activator module of the reference parameter set are listed in Table 2.5.

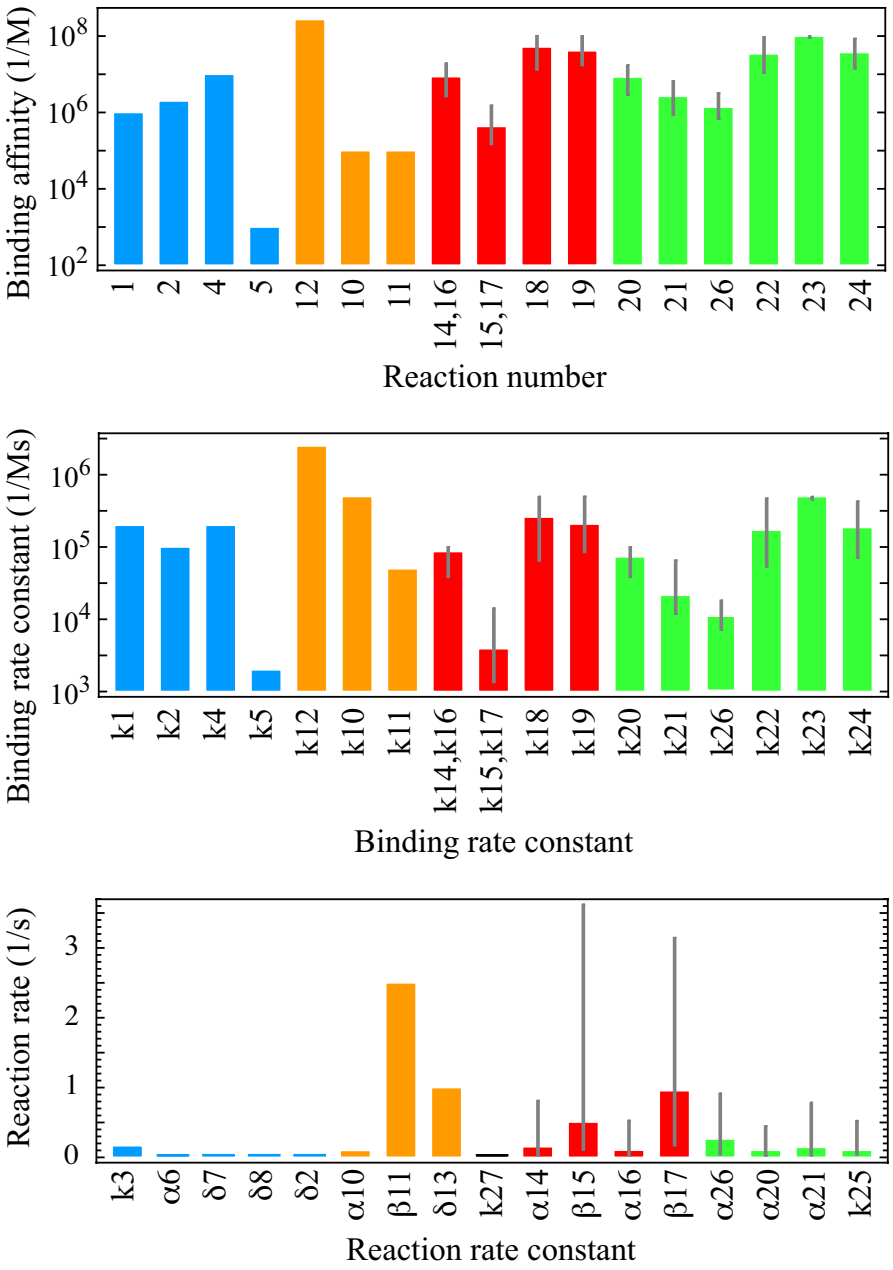


Figure 2.5: The optimization of the systems functionality resulted in distinct values for most kinetic rate constants. The parameters in the licensing module (colored in blue) and the S-Cdk activation module (orange) are fixed in all parameter sets. Median values and the 1/6- and 5/6-quantile deviations are shown for the kinetic rate constants of the 11-3-2 activator module (red) and the firing module (green) in all optimized parameter sets.

Table 2.5: Optimized values for the kinetic rate constants in the firing and 11-3-2 activator module listed for the reference parameter set.

Parameter	Symbol	Value	Unit	Remarks
11-3-2 activator formation				
Binding of S-Cdk and Sld2/3	k_{14}, k_{16}	6.410×10^4	$(\mathrm{Ms})^{-1}$	
Dissociation of S-Cdk and Sld2/3	k_{-14}, k_{-16}	0.008	s^{-1}	
Phosphorylation of Sld2/3	α_{14}	0.223	s^{-1}	
Phosphorylation of Thr84 in Sld2	α_{16}	0.945	s^{-1}	
Binding of Cdc14 and Sld2/3	k_{15}, k_{17}	1.146×10^3	$(\mathrm{Ms})^{-1}$	
Dissociation of Cdc14 and Sld2/3	k_{-15}, k_{-17}	0.010	s^{-1}	
Dephosphorylation of Sld2/3	β_{15}	0.276	s^{-1}	
Dephosphorylation of Thr84 in Sld2	β_{17}	0.460	s^{-1}	
Binding of Sld2p and Dpb11	k_{18}	5.000×10^5	$(\mathrm{Ms})^{-1}$	
Dissociation of Sld2p and Dpb11	k_{-18}	0.005	s^{-1}	fixed
Binding of Sld3p and Dpb11	k_{19}	1.709×10^5	$(\mathrm{Ms})^{-1}$	
Dissociation of Sld3p and Dpb11	k_{-19}	0.005	s^{-1}	fixed
Firing				
Binding of S-Cdk and Mcm2-7	k_{20}	7.310×10^4	$(\mathrm{Ms})^{-1}$	
Dissociation of S-Cdk and Mcm2-7	k_{-20}	0.008	s^{-1}	
Phosphorylation of Mcm2-7 by S-Cdk	α_{20}	0.141	s^{-1}	
Binding of Ddk and Mcm2-7	k_{21}	2.977×10^4	$(\mathrm{Ms})^{-1}$	
Dissociation of Ddk and Mcm2-7	k_{-21}	0.015	s^{-1}	
Phosphorylation of Mcm2-7 by Ddk	α_{21}	0.790	s^{-1}	
Binding of Cdc45	k_{22}	1.087×10^5	$(\mathrm{Ms})^{-1}$	
Dissociation of Cdc45	k_{-22}	0.005	s^{-1}	fixed
Binding of Sld3p-Dpb11-Sld2p	k_{23}	5.000×10^5	$(\mathrm{Ms})^{-1}$	
Dissociation of Sld3p-Dpb11-Sld2p	k_{-23}	0.005	s^{-1}	fixed
Binding of GINS–DNA polymerase	k_{24}	3.310×10^5	$(\mathrm{Ms})^{-1}$	
Dissociation of GINS–DNA polymerase	k_{-24}	0.005	s^{-1}	fixed
Dissociation of 11-3-2 activator	k_{25}	0.927	s^{-1}	
Binding of S-Cdk and Orc subunits	k_{26}	6.325×10^3	$(\mathrm{Ms})^{-1}$	
Dissociation of S-Cdk and Orc subunits	k_{-26}	0.009	s^{-1}	
Phosphorylation of Orc subunits	α_{26}	0.676	s^{-1}	
Dissociation of Mcm2-7, Cdc45, and GINS–DNA polymerase / resetting of origins after replicon completion	k_{27}	0.002	s^{-1}	fixed

3 Analysis of the kinetics of DNA replication initiation

In this Chapter, the functionality of the mathematical model parameterized with the different parameter sets is quantified and analyzed in comparison to experimental observations. The dependency of the functional systems properties on the kinetic parameters and the initial protein concentrations is calculated to estimate how these control the performance of the network in terms of rate and coherence of origin activation, number of activated origins, and robustness against DNA rereplication. Finally, characteristic perturbations of the S-Cdk activity, which were also studied experimentally, are rationalized using the mathematical model.

3.1 Origin firing kinetics with functional parameter sets

The firing rates, $f(t)$, generated with the mathematical model with different optimized parameterizations show that in the majority of parameter sets (65%) origin firing starts between 35 to 38 minutes after the beginning of origin licensing and is completed in less than 15 minutes (Figure 3.1(a)). For some parameter sets, the firing rate reaches values above one origin firing event per second, implying a very synchronous activation of replication origins. Only a few parameter sets (11%) were admitted that take longer than 30 minutes to finish origin firing.

The experimental determination of profiles of DNA replication in the budding yeast genome are based on DNA microarray technology [50]. Yabuki et al. (2002) [88] monitored the change of the DNA copy number from one to two during DNA replication using flow cytometry, specifically fluorescence activated cell sorting (FACS). The DNA content was determined in 23 samples, collected every 2.5 minutes during S phase of a synchronized cell population that was previously released from α-factor, a peptide that is widely used to arrest the cell cycle in G1 phase [8]. At the same time, genomic DNA was isolated, enzymatically fragmented into pieces of $\sim$ 1.5 kb on average, labeled at the 3' ends, and finally hybridized to high-density oligonucleotide microarrays. By this method, the distribution of firing times of $\sim$ 250 replication origins was monitored. The activation times of only the early replicating origins were determined by the addition of hydroxyurea (HU) to the cell population, which is supposed to block the initiation of late origins (Figure 3.1(b)). The activation of the $\sim$ 114 early origins was reported to start 17 minutes after release from α-factor and to be completed after 14 minutes, which corresponds to a duration of approximately 23 minutes for the activation of a total number of 190 early origins.

Another study by Raghuraman et al. (2001) [67] uses isotopic labeling of DNA to distinguish between newly synthesized and previously existing DNA. Two cell cycle blocks, α-factor and a temperature sensitive Cdc7, which is the catalytic subunit of the Ddk kinase, were used to

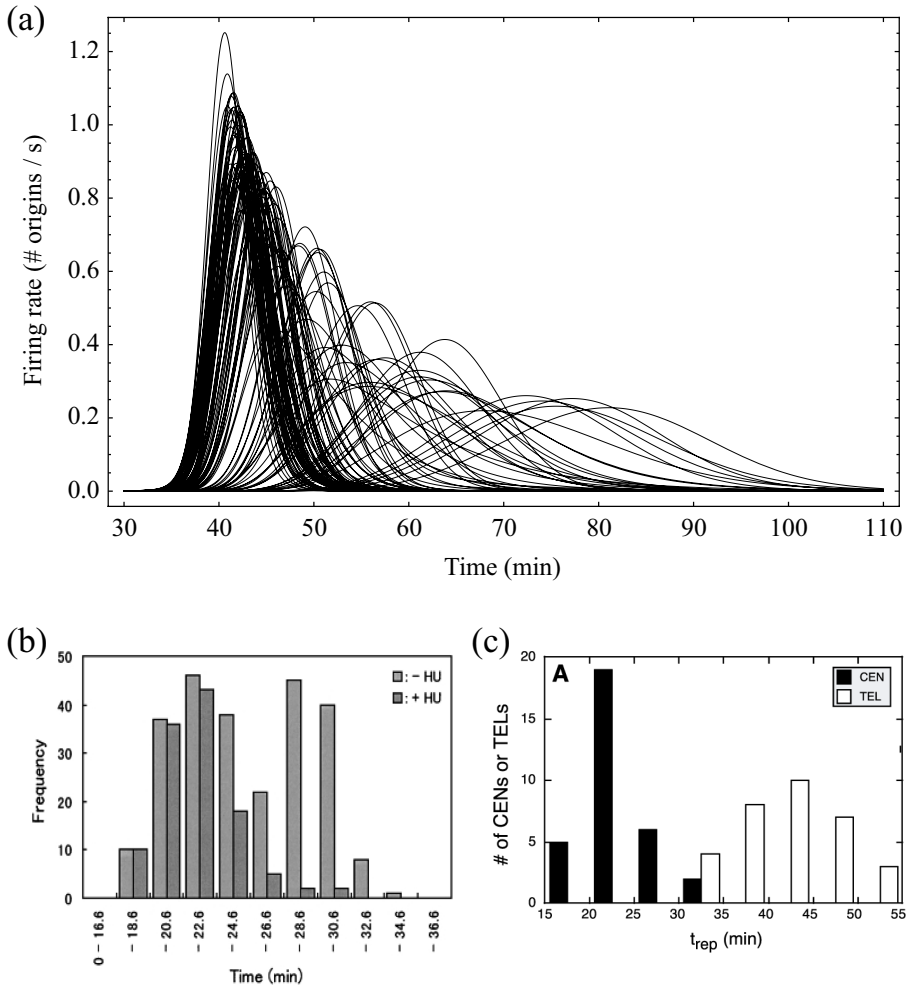


Figure 3.1: (a) Firing rates, $f(t)$, generated with the 109 optimized parameter sets exhibit some variations in the functional kinetics of origin firing. (b) Yabuki et al. (2002) [88] determined the firing times of all replication origins (light grey) and of only the early origins (dark grey). (c) Raghuraman et al. (2001) [67] monitored the time of replication of centromeres and telomeres.

synchronize and arrest cells that were grown in heavy medium (^{13}C and ^{15}N) for several generations, before the onset of DNA replication in S phase. The medium was then changed from heavy to light medium (^{12}C and ^{14}N) and cells were released again from the cell cycle blocks. In this way, all previously existing DNA consists of a heavy DNA double strand, whereas all newly synthesized DNA comprises one light and one heavy DNA single strand. Potentially rereplicated DNA can be identified by one DNA double strand that is only made up of light nucleotides. At eight time points within one hour during S phase, samples of genomic DNA were collected. Enzymatically fragmented DNA pieces of ~ 50 bp on average are separated into

two fractions of the different density labels by density-gradient centrifugation. Finally, samples were end-labelled and hybridized to oligonucleotide microarrays. In this way, the replication dynamics in the yeast genome was monitored. Replication of the 32 centromeric regions that belong to the early replicating part of each chromosome was observed to start 15 minutes after the release from α-factor and was finished within approximately 20 minutes (Figure 3.1(c)).

The kinetics of origin firing simulated with the mathematical model is, in most cases, consistent with the experimental observations. In particular, the replication kinetics generated with some faster parameter sets agrees with the reported replication times of the early replication origins.

3.1.1 Functional kinetics in the different parameter sets

All admissible parameter sets generate a kinetics of origin firing, that account for the required functionality of the system. The mathematical formulation of the functional systems properties allows to quantify and compare the functionality of the firing kinetics with the different parameter sets.

The distribution of the values of the functional system properties in the generated parameter sets (Figure 3.2) shows that for most parameter sets the number of rereplicating origins, ρ, as well as the mean time to firing, τ, and the duration of firing, Δ, attain a relatively small value of $\rho < 0.0015$ origins, $\tau < 46$ minutes, and $\Delta < 8$ minutes, respectively. The achieved number of firing origins, N, is approximately equally distributed between 181 and 185 origins in the generated parameter sets. The mean time to firing, τ, as well as the duration of firing, Δ, seem to be bordered towards short times. No parameter sets are generated that exhibit a mean time to firing, τ, of less than 40 minutes and a duration of firing, Δ, shorter than 4 minutes, although no constraint for these times was specified in the optimization process.

As the appearance of the firing rates, $f(t)$, suggests (Figure 3.1(a)), a correlation exists between the mean time to firing, τ, and the duration of firing, Δ (Figure 3.3(a)). "Faster" parameter sets that lead to an early initiation of firing at the same time tend to exhibit a short duration of origin firing, Δ, whereas late firing parameter sets also give rise to a temporally spread initiation of origin firing. In particular the faster parameter sets agree with the observed narrow distribution of firing times of the early origins (Figure 3.1(b)).

Interestingly, a strong non-linear correlation is found for the timing of origin firing, characterized by τ and Δ, and the probability for the occurrence of DNA rereplication, ρ (Figure 3.3(b), not shown for Δ and ρ). Parameter sets that result in an early activation of origins and a short duration of firing tend to exhibit a higher probability for rereplication, ρ, than those parameter sets that lead to a later initiation of origin firing. Hence, a rapid progression to and activation of origin firing appears to have negative effects on the proper inhibition of DNA rereplication. Parameter sets can be found, in which origin firing is even faster than 40 minutes, but for all of these the number of rereplicating origins exceeds the assumed limit of $\rho = 0.01$ origins (not shown).

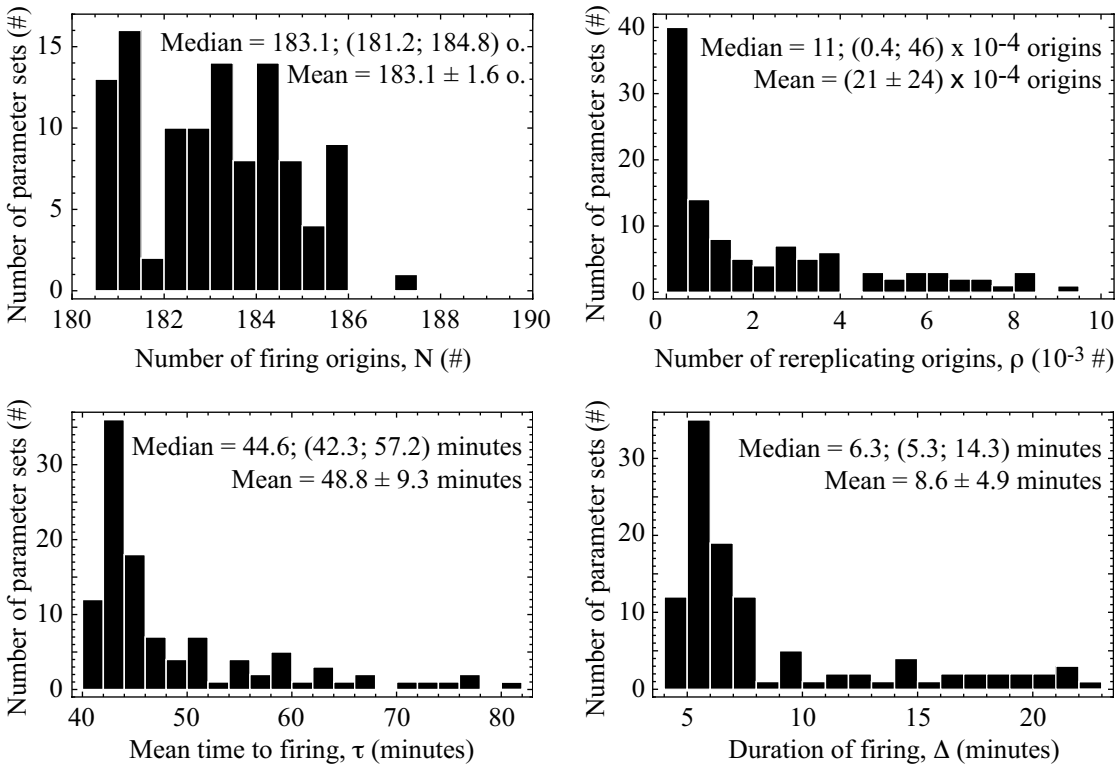


Figure 3.2: The values of functional systems properties, N, ρ, τ, and Δ, generated by the optimized parameter sets exhibit a distribution. While the number of firing origins, N, is equally distributed between 181 and 185 origins, the distribution of the number of rereplicating origins, ρ, the mean time to origin firing, τ, and the duration of firing, Δ, appears to be log-normal and is better characterized by the median value and the 1/6- and 5/6-quantile deviations.

3.1.2 Dependency on the initial number of early origins

The precise number of early origins that fire in one S phase and their composition is not well-defined experimentally, and it may not be the same in every cell and each cell cycle. In the mathematical model, we referred to an initial number of 190 early origins, which was identified experimentally by Lengronne et al. (2001) using hydroxyurea (HU) to block the initiation of late firing origins [46]. Furthermore, we assumed that almost all of these early origins (95%) become active in each cell cycle. The identity of origins is not an object of the mathematical model, thus, the composition of the early origins that finally initiate DNA replication is not necessarily the same in every cell cycle.

In order to exclude, that the functionality of the replication initiation network is dependent on or biased by the initial number of early origins put into the mathematical model, the functional system properties are calculated for an initial number of early origins varying between 50 to 800

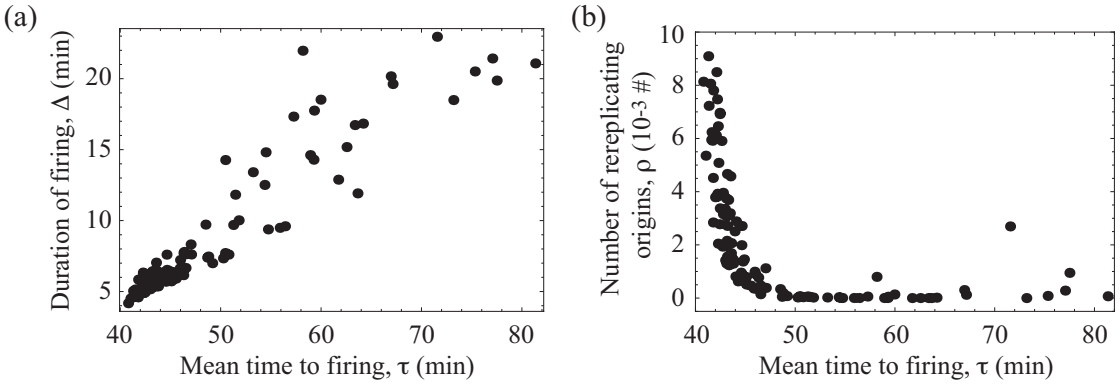


Figure 3.3: (a) A correlation between the mean time to origin firing, τ, and the duration of firing, Δ, is identifiable in the generated parameter sets (correlation coefficient: 0.94). (b) The mean time to firing, τ, and the logarithm of the number of rereplicating origins, ρ, are anti-correlated (correlation coefficient: -0.67). Δ and ρ exhibit a similar relationship (correlation coefficient: -0.58; not shown).

origins, simulated with the reference parameter set (Figure 3.4). The ratio between the initial number of early origins and the number of origins that are eventually activated and initiate DNA replication, remains constant. The number of rereplicating origins, ρ, the mean time to firing, τ, and duration of firing, Δ, also attain roughly the same value within this range of 50 to 800 initial early origins.

Thus, the functional systems properties do not depend on the initial number of early origins, and would even be fulfilled if the 730 autonomously replicating sequences (ARS) that are up to now identified as potential replication origins in the budding yeast genome [60] would all and simultaneously be activated early in the cell cycle.

3.2 Control of functional systems properties

A well-established method to analyze the sensitivity of a system towards the change in a particular parameter value is the calculation of sensitivity coefficients [33]. The sensitivity coefficient gives the relation between the relative change in a systems property in response to the relative change of a particular parameter value. It is defined as follows:

$$C_i^{\xi} = \frac{\partial \ln \xi}{\partial \ln p_i} = \frac{\partial \xi}{\partial p_i} \cdot \frac{p_i}{\xi} = \frac{\partial \xi}{\xi} \left(\frac{\partial p_i}{p_i} \right)^{-1},$$

where ξ is the systems property to be analyzed (in this case N, ρ, τ and Δ), and p_i the varied parameter value (in this case the concentration or kinetic rate constant).
A positive sensitivity coefficient implies an increase of the systems property, whereas a negative coefficient describes the decline of the systems property in response to an elevation of a

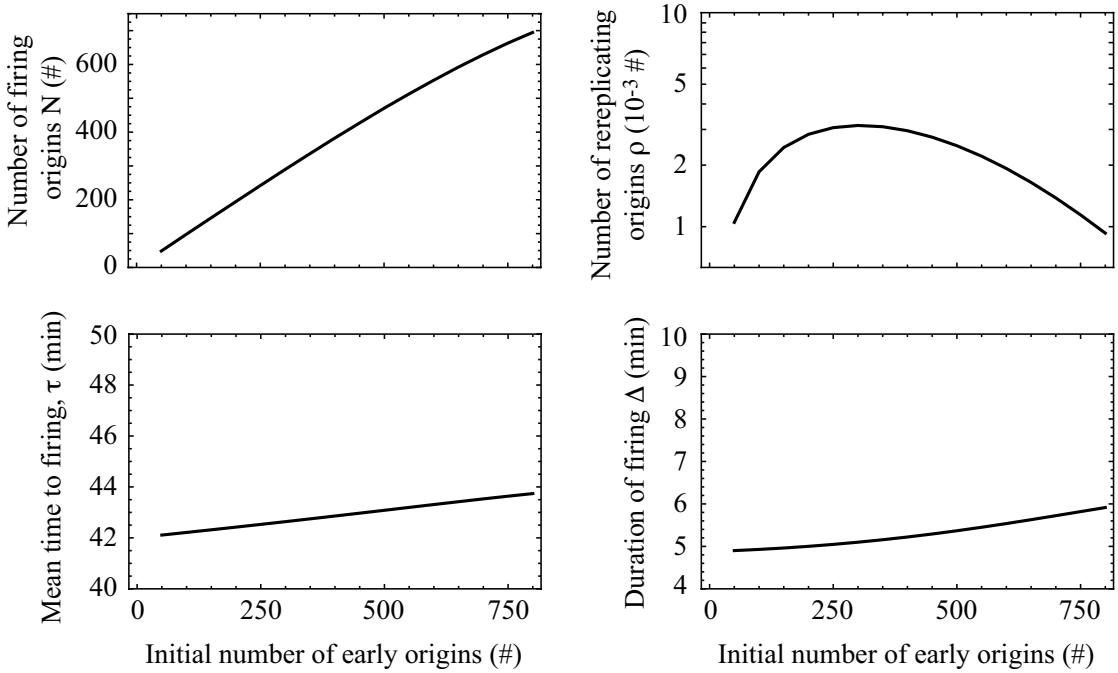


Figure 3.4: The functional systems properties, N, ρ, τ, and Δ, of the kinetics of origin firing are not affected by the number of initial early origins considered in the mathematical model, in the range between 50 and 800 initial early origins.

particular parameter value, respectively.

3.2.1 Control by protein concentrations

The protein concentrations in a living cell are subject to natural fluctuations, e.g. gene expression is a stochastic process and cell and nuclear volumes vary from cell to cell and during cell growth [42]. To analyze the effect of changes in the protein concentrations on the kinetics of origin firing, the concentration sensitivity coefficients are calculated for all four functional systems properties (N, ρ, τ and Δ) in response to variations in all 13 initial molecule concentrations (see Table 2.2). The concentration sensitivity coefficients are determined for each generated parameter set resulting in a set of sensitivity coefficients with different values originating from each parameter set. The consideration of all generated, functional parameter sets allows to assess the different values obtained for a sensitivity coefficients of a functional systems property. In this way, characteristic controls are identified and distinguished from undefined sensitivity contributions, which only appear in very few generated parameter sets.

The results for the concentration sensitivity coefficients are summarized in the median and 1/6-, 5/6-quantile values of the sensitivity coefficients calculated with all parameter sets (Figure 3.5).

The amplitudes of the sensitivity coefficients of the different systems properties reveal that

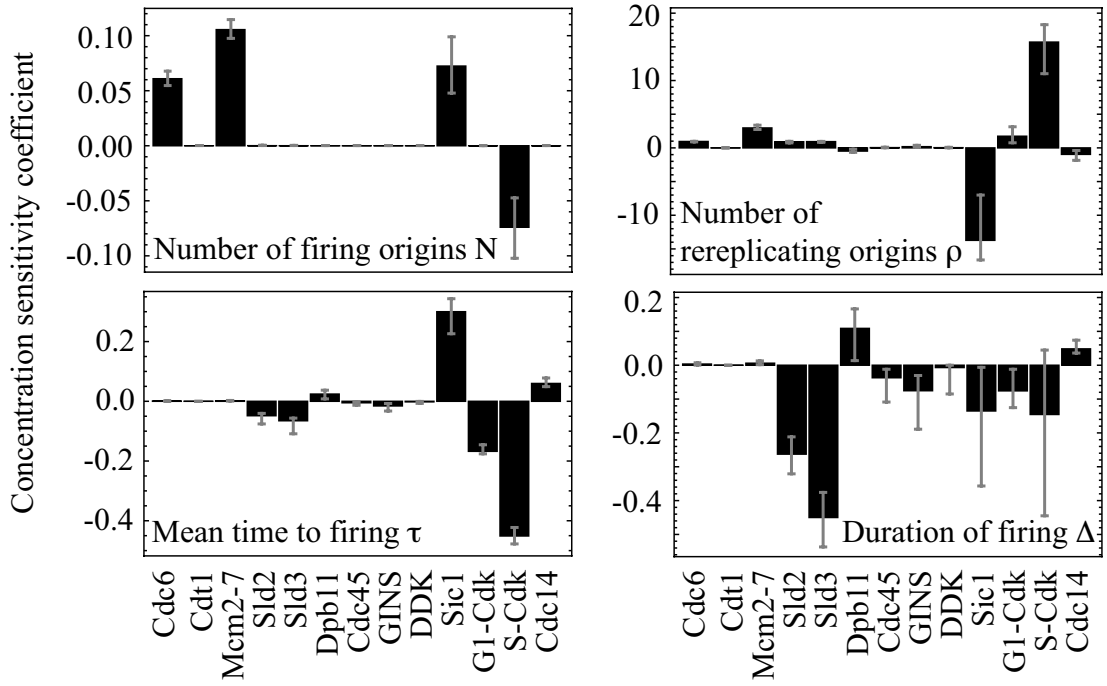


Figure 3.5: The concentration sensitivity coefficients quantify the dependency of the functional systems properties, N, ρ, τ and Δ, on the initial amount of each molecule considered in the mathematical model. Shown are the median, 1/6-quantile, and 5/6-quantile values of the concentration sensitivity coefficients calculated with all optimized parameter sets.

the number of firing origins, N, the mean time to firing, τ, and the duration of firing, Δ, are robust systems properties, with concentration sensitivity coefficients of maximally 0.5, while the number of rereplicating origins, ρ, appears to be a very sensitive property with concentration sensitivity coefficients of up to around 18.

The number of firing origins, N, is only dependent on the concentrations of the licensing factors, Cdc6 and Mcm2-7, and on the amount of Sic1 and S-Cdk, that are important for the inhibition of S-Cdk activity at the beginning of the licensing phase. The number of rereplicating origins, ρ, is mainly controlled by the amount of Sic1 and S-Cdk molecules that regulate the activation of S-Cdk, and thereby the initiation of S-Cdk-dependent mechanisms to prevent DNA rereplication. Dependencies on the concentration of the licensing factors, Cdc6 and Mcm2-7, the 11-3-2 activator constituents, Sld2, Sld3 and Dpb11 and the enzymes involved in the activation of S-Cdk, that are G1-Cdk and Cdc14, are also identifiable. The mean time to firing, τ, is primarily influenced by proteins involved in the release of active S-Cdk and the phosphorylation of Sld2 and Sld3, which are Sic1, G1-Cdk, S-Cdk and Cdc14. The duration of origin firing, Δ, however, is mainly affected by fluctuations in the concentrations of the 11-3-2 activator constituents, Sld2, Dpb11 and, particularly, Sld3, but also the amount of Sic1 and S-Cdk molecules,

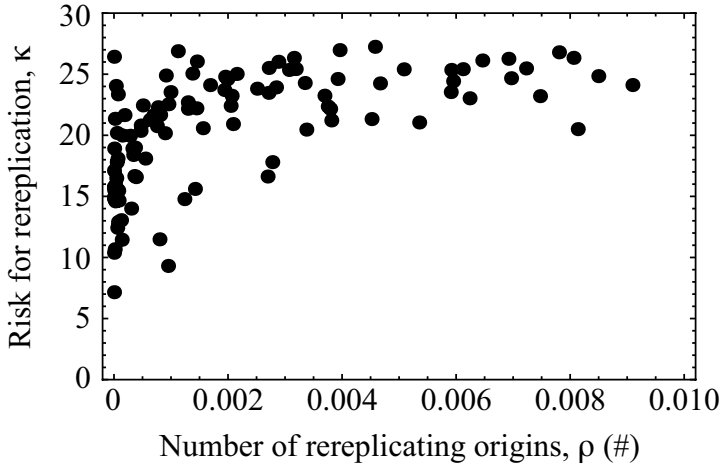


Figure 3.6: The number of rereplicating origins, ρ, and the risk for rereplication, κ, are correlated, indicating that parameter sets causing a higher number of rereplicating origins simultaneously carry a higher risk for rereplication.

regulating the release of active S-Cdk, control this property. Interestingly, an increase in the concentration of Sic1 prolongs the mean time to firing, τ, and, at the same time, causes a shortening of the duration of firing, Δ.

Summarizing, the activation and the activity of S-Cdk exert a strong control on all functional systems properties, as it is involved in both, initiation of DNA replication and inhibition of DNA rereplication.

The overall concentration sensitivity coefficient of a parameter set, referred to as the risk for rereplication, κ, is defined as the sum in quadrature (Euclidian norm) of all 13 concentration sensitivity coefficients for the number of rereplicating origins, $C^{\rho}_{p_i}$,

$$\kappa = \left(\sum_{i=1}^{13} \left(C^{\rho}_{p_i} \right)^2 \right)^{1/2}.$$

Remarkably, a correlation is found for the number of rereplicating origins, ρ, and the risk for rereplication, κ (Figure 3.6). Optimized functional parameter sets that give rise to a relatively high number of rereplicating origins, ρ, that is to a high probability for the occurrence of rereplication, additionally, tend to be most sensitive towards fluctuations in the protein concentrations. And parameter sets that generate a very low number of rereplicating origins are, at the same time, very robust towards the emergence of rereplication.

3.2.2 Control by kinetic parameters

To analyze the effect of a particular kinetic parameter on the functionality of the replication initiation network, parameter sensitivity coefficients are calculated for all four functional systems properties in response to changes in the 51 kinetic parameters. Although the parameters of the licensing and S-Cdk activation module are fixed in all admissible parameter sets, variations in the sensitivity coefficients of these parameters occur between the parameter sets, because the different parameter values in the firing and 11-3-2 activator module modify the sensitivity contribution of the fixed parameters in each parameter set.

The sensitivity coefficients are calculated for each optimized parameter set. The results of the sensitivity analysis for all four systems properties in response to changes in the kinetic parameters is again summarized in the median and 1/6-, 5/6-quantile values of all sensitivity coefficients (Figure 3.7).

The observations made in the analysis of the concentration sensitivity coefficients are confirmed by the analysis of the parameter sensitivity coefficients, but latter coefficients additionally allow to relate the total control of a molecule to the partial control contributions from the reactions, in which the molecule is involved.

As already identified, the number of rereplicating origins, ρ, appears to be a very sensitive systems property. Variations in the kinetic rate constants, especially changes of parameters in the licensing and S-Cdk activation modules, result in parameter sensitivity coefficients of up to 6, while the number of firing origins, N, the mean time to firing, τ, and the duration of firing, Δ, are robust systems properties with maximal sensitivity coefficients of around 0.6.

Licensing parameters only play a role in the regulation of the number of firing origins, N, and the number of rereplicating origins, ρ; both critical systems properties presumably implicated in the emergence of genomic instability.

In addition to the strong sensitivity to most licensing parameters, the number of firing origins, N, is controlled by reactions determining the inhibition of active S-Cdk during the licensing phase, such as the affinity of the stoichiometric inhibitor Sic1 to S-Cdk, and by reactions causing the inhibition of DNA (re)licensing, precisely, the binding rate of S-Cdk to Orc molecules leading to the phosphorylation of Orc subunits. The number of rereplicating origins, ρ, is affected by almost all parameters of the licensing and S-Cdk activator module. The strongest control is exerted by the binding rate of S-Cdk to Sld2 and Sld3, which leads to an increased number of rereplicating origins, ρ, at a faster reaction rate constant. The binding rate constant of the 11-3-2 activator to the origins and the reaction rate constant describing the transition between the firing of an origin and its reentering at the initial origin state, $S0$, in the licensing module, have a minor influence on the number of rereplicating origins, ρ. As presumed in the optimization process, the timing of origin firing, described by the mean time to firing, τ, and the duration of firing, Δ, is independent of kinetic parameters in the licensing module. It is controlled by the reactions describing the release of active S-Cdk from inhibition by Sic1, such as the affinity of Sic1 to S-Cdk and the affinity and phosphorylation rate of G1-Cdk to Sic1. Reactions determining the formation and the association of the 11-3-2 activator complex with the origins, in particular, the binding rate of S-Cdk to Sld2 and Sld3 and the binding rate of the 11-3-2 activator to the origins, strongly influence the origin firing kinetics, τ and Δ. A higher affinity of Sld2 and Sld3 to Dpb11 accelerates the duration of firing, Δ, but hardly influences the mean time to firing, τ.

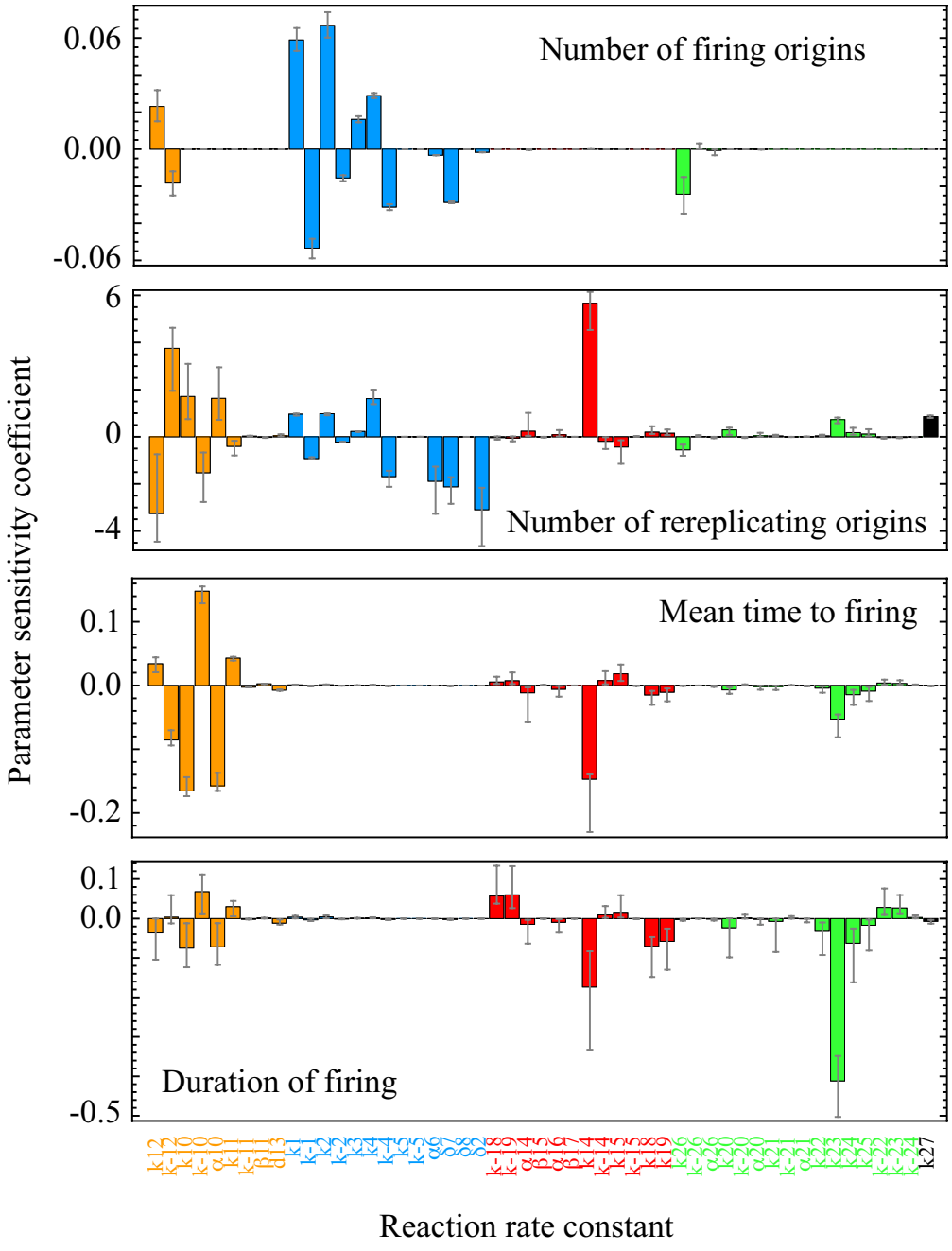


Figure 3.7: The parameter sensitivity coefficients quantify the dependence of the functional systems properties, N, ρ, τ and Δ, on the values of the kinetic rate constants. Shown are the median, 1/6-quantile, and 5/6-quantile values of the parameter sensitivity coefficients calculated with all optimized parameter sets.

3.2.3 Control by S-Cdk concentration

The sensitivity coefficients of both, the protein concentrations and the kinetic parameters, reveal a pivotal role for S-Cdk in the regulation of the systems performance. As a kinase, S-Cdk catalyzes activating protein phosphorylations of Sld2, Sld3, and Mcm2-7 molecules at the origins, as well as in DNA rereplication inhibiting mechanisms through the phosphorylation of Orc subunits and free Mcm2-7 molecules. Consequently, the release of active S-Cdk has to be tightly controlled by Sic1 to ensure a period of low S-Cdk activity, in which origins are licensed. The impact of perturbations in the S-Cdk activation kinetics on the kinetics of origin firing has been studied experimentally and is rationalized in the last section of this chapter.

To analyze the dependency of the functional kinetics of origin firing on the initial number of S-Cdk molecules, the values of the four functional systems properties are calculated for initial S-Cdk concentrations ranging from 100 to 1300 molecules per cell, which corresponds to a decrease of the reference value for the initial number of S-Cdk molecules (600) to one sixth and a roughly doubling of it. The resulting values of the functional systems properties are summarized in the median values and 1/6-, 5/6-quantile deviations of all generated parameter sets (Figure 3.8).

All systems properties are strongly influenced by a change in the S-Cdk concentration. The number of rereplicating origins, ρ, is even varied over several orders of magnitude. The number of firing origins, N, and the number of rereplicating origins, ρ, would benefit from a low number of S-Cdk molecules, since, in this case, the number of firing origins,N, is as high as possible and, at the same time, the number of rereplication origins, ρ, is low as requested. A high number of S-Cdk molecules causes a reduced number of firing origins, N, possibly because of a too strong inhibition of origin (re)licensing through the phosphorylation of Orc subunits by S-Cdk. At the same time, a high concentration of S-Cdk leads to an increase of the number of rereplicating origins, ρ, which could be a side effect of a too fast activation of replication origins, after which the mechanisms inhibiting DNA rereplication, such as degradation of Cdc6, export of Mcm2-7 and Cdt1, are not then fully operative. However, for almost all parameter sets, the maximum number of rereplicating origins, ρ, stays well below 10 origins and for most optimized parameter sets even below one rereplicating origin. The timing of origin firing would be fastest and most synchronous for a high concentration of S-Cdk, while at low S-Cdk concentrations, the origin firing would initiate delayed and take much longer. With less than 200 S-Cdk molecules per cell, the mean time to firing, τ, could become even longer than the average duration of a cell cycle in budding yeast, which takes about two hours [64]. Compared to this, the duration of firing, Δ, is prolonged only moderately at very low S-Cdk concentrations. At concentrations between ~ 200 and 700 S-Cdk molecules per cell, the duration of firing, Δ, appears to be almost constant.

According to this analysis, the kinetics of origin firing would be most favorable, that is giving rise to a high number of firing origins, N, a low number of rereplicating origins, ρ, and a reasonably fast timing of origin firing, τ and Δ, at an intermediate concentration of S-Cdk, as is indeed observed experimentally [27] and used here in the mathematical model.

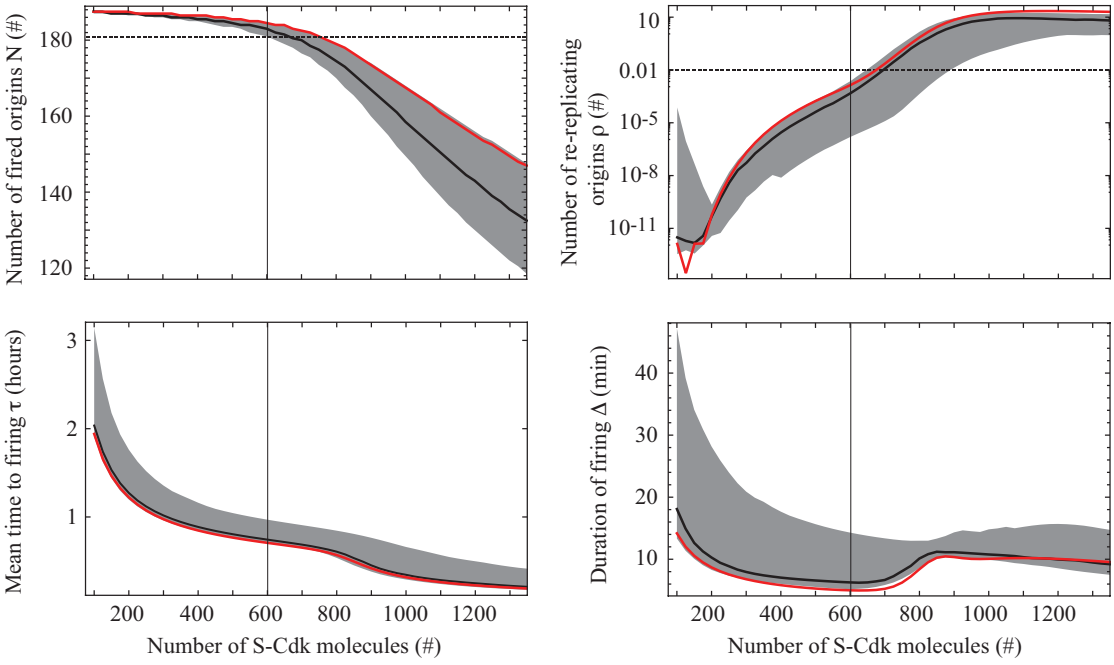


Figure 3.8: The functional systems properties are strongly influences by the initial number of S-Cdk molecules (red line: reference parameter set; black line: median of all optimized parameter sets; grey area: 1/6- and 5/6-quantile deviations from the median value), but a compromise between rapid and accurate initiation of DNA replication can is found for a range of medium S-Cdk concentrations, from approximately 200 to 700 molecules per cell.

3.2.4 Control by multisite phosphorylation of Sld2 and Sic1

Sic1 and Sld2 are both phosphorylated at multiple residues by G1-Cdk and S-Cdk, respectively, which causes a time delay between the activation of G1-Cdk and S-Cdk, in the case of Sic1, and between the activation of S-Cdk and the initiation of origin activation, in the case of Sld2. At the same time, random multisite phosphorylation of a substrate contributes to its sharp transition to the fully phosphorylated state. In the following, the specific function of each time delay is analyzed by varying the number of random phosphorylation events in Sld2 and Sic1.

In the case of Sld2, the number of random phosphorylations, occurring before the essential, sequential Thr84 phosphorylation, is changed from zero to nine hypothetical random phosphorylations in Sld2 by S-Cdk (Figure 3.9, left column). In the case of Sic1, the minimum required number of phosphorylations before the recognition and degradation of Sic1 is varied from one to nine, while the total number of nine random phosphorylations in Sic1 by G1-Cdk remains constant (Figure 3.9, right column).

Both, multisite phosphorylation of Sld2 and Sic1, have no influence on the number of fired origins, N, indicating that this number is already fixed during the licensing phase and indepen-

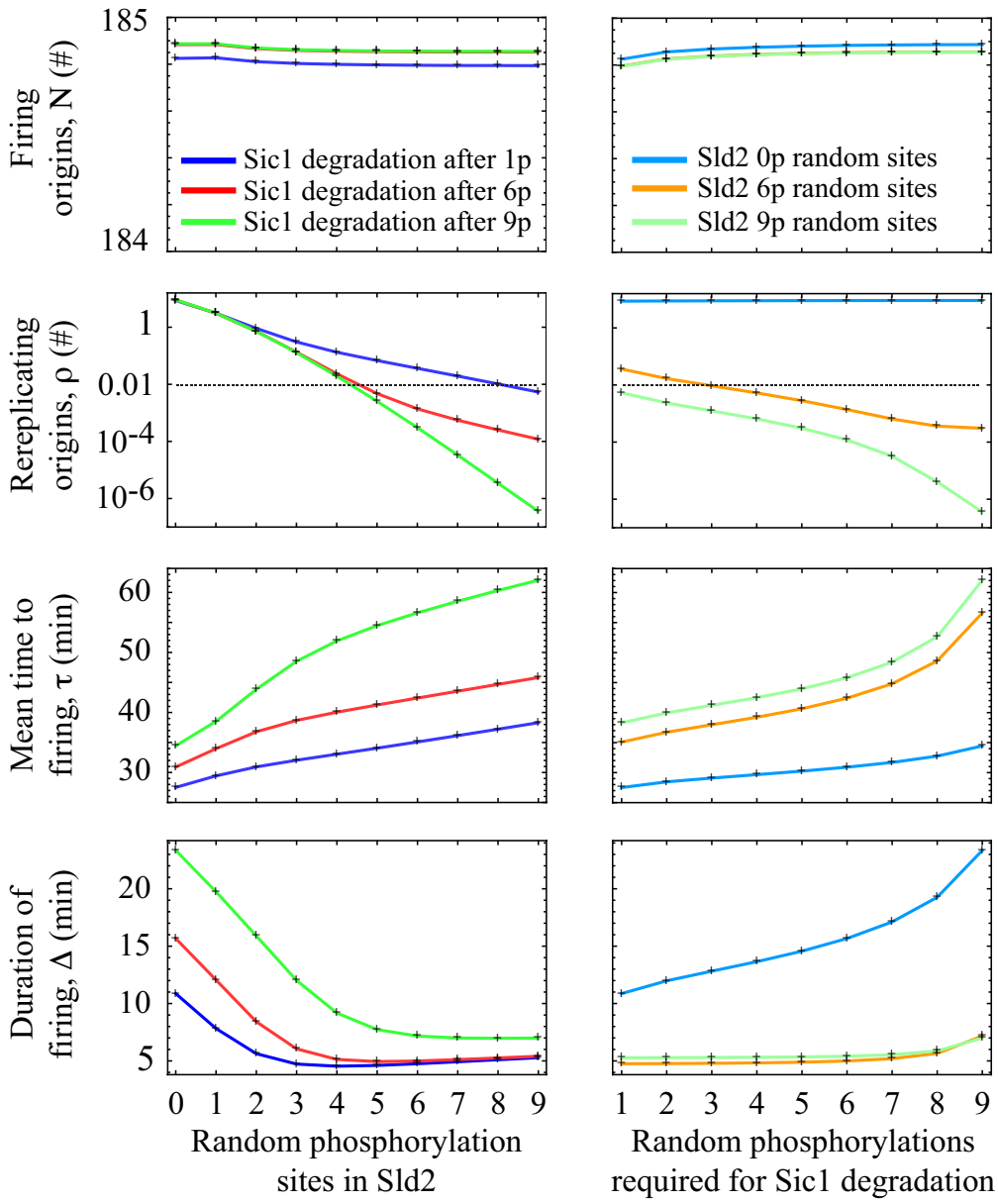


Figure 3.9: The dependency of the functional systems properties on the number of random phosphorylation events in Sld2 and Sic1 by S-Cdk and G1-Cdk, respectively, is realized by a reduction of the number of random phosphorylations before the threonine 84 phosphorylation in Sld2 and by varying the number of random phosphorylations that are required for the degradation of Sic1, while maintaining the total number of random phosphorylations in Sic1. The reference parameter set and mathematical models accordingly modified are used to calculate the functional systems properties.

dent of processes outside the licensing phase.

The number of rereplicating origins, ρ, on the contrary, is strongly affected by the number of random phosphorylation events in Sld2 and Sic1. A higher number of phosphorylation events in each molecule increases the time delay between origin licensing and firing that is needed for rereplication inhibiting mechanisms to set in, and thereby reduces the probability for DNA rereplication. Variation of the number of phosphorylation sites in Sld2 reveal that between five and six random phosphorylations are needed to fulfill the functional requirements of less that 0.01 rereplicating origins, ρ, per cell cycle. The same behavior is identifiable, when additionally the Sic1 protein is hypothetically degraded not before nine phosphorylations. A premature degradation of Sic1 after already one phosphorylation only results in a functional kinetics of origin firing, when simoultaneously the number of random phosphorylation events in Sld2 are assumed to be more than eight. Variations of the number of phosphorylations necessary for Sic1 degradation show that actually only four phosphorylations in Sic1 before its degradation are needed to fulfill the functional requirement of less than 0.01 rereplicating origins, ρ, per cell cycle. A number of less than four phosphorylations required for the degradation of Sic1 would only lead to a functional kinetics of origin firing, if additionally the Cdc4-mediated recognition of the phosphorylated Sic1 protein is assumed to be slower (not shown), which would be described by a smaller rate constant for the Sic1 degradation in the mathematical model. The exact value for the degradation rate constant of Sic1 has not been determined experimentally.
An already low number of rereplicating origins, ρ, caused by a high number of phosphorylation events in Sld2 (nine phosphorylations in Figure 3.9) could be reduced additionally by a high number of phosphorylations required for Sic1 degradation, whereas an insufficient Sld2-dependent time delay, due to the absence of the multiple random phosphorylation in Sld2, results in a number of around 10 rereplicating origins, ρ, which can not be rescued by a prolonged Sic1-dependent time delay.

The mean time to firing, τ, is prolonged similarly in Sld2 and Sic1, by an increased number of phosphorylation events. Interestingly, the steepest increase in the mean time to firing, τ, is observed at a very high number of phosphorylation sites in the case of Sic1 and at a very low number of phosphorylation sites in the case of Sld2.

The multisite phosphorylations of Sld2 and Sic1 have opposing effects on the duration of firing, Δ. The duration of firing, Δ, increases strongly with a number of less than four random phosphorylation events in Sld2. A minimal duration of firing, Δ, is found for a number of approximately five random phosphorylation events in Sld2. This is not very surprising, since the kinetic parameter values were determined by minimizing the duration of firing, Δ, at a given structure of the molecular network with six random phosphorylations in Sld2 and one sequential Thr84 phosphorylation. For a higher number of random phosphorylations in Sld2, the duration of firing, Δ, remains approximately constant. The duration of firing, Δ, is almost independent of the multisite phosphorylation of Sic1. Only a number of more than seven random phosphorylations in Sic1 needed for the degradation of Sic1, could lead to a slightly prolonged duration of firing, Δ. If the random phosphorylation in Sld2 is not existent, the duration of firing, Δ, is increased considerably and even rises with an additional number of phosphorylations in Sic1.

The analysis of the multiple phosphorylation in Sld2 and Sic1 confirms the importance of both proteins in creating a time delay. The Sic1-dependent time delay occurs before the activation of S-Cdk, whereas Sld2 creates a time delay before the initiation of origin firing. According to

this analysis, the specific number of random phosphorylation sites in Sld2 appears to be more important for the functional kinetics of origin firing. The random multisite phosphorylation of Sld2 by S-Cdk results in a deferred formation of the 11-3-2 activator complex and with this in a delayed initiation of origin firing. This leads to a temporal separation between the activating and the inhibiting functions of S-Cdk.

3.3 Simulations by the mathematical model

In the following, the mathematical model is used to simulate the initiation of DNA replication under normal conditions and for characteristic perturbations of the regulation of S-Cdk activity, which also has been studied experimentally. The functionality of the replication initiation network is quantified by the functional systems properties allowing to identify typical differences in the firing kinetics under normal and perturbed conditions. The consequences of an impaired kinetics of origin firing on the processes happening in S phase are analyzed in Chapter 4.

3.3.1 Model simulation under normal conditions

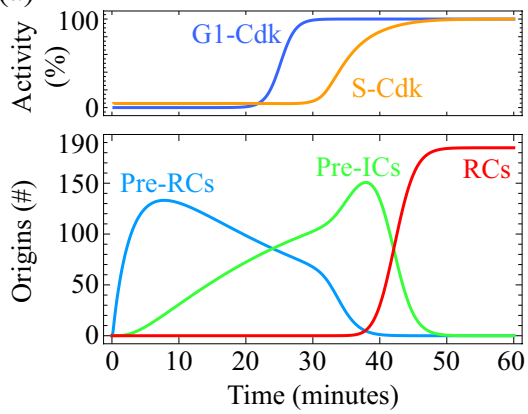
The kinetic parameters of the mathematical model were optimized, such that all functional requirements are fulfilled under normal conditions. From all 109 optimized parameter sets, one parameter set was selected as a reference parameter set, which is among the 11% with the shortest duration, Δ, the 20% with the fastest onset of firing, τ, and, at the same time, the 70% with the lowest probability of rereplication, ρ, and the 16% with the most firing origins, N. The values of the kinetic rate constants of the reference parameter set are listed in Table 2.3 and 2.5.

Under normal conditions, S-Cdk becomes active approximately ten minutes after the activation of G1-Cdk (Figure 3.10(a), upper panel), which is described by a sigmoid function (equation (2.16) in the mathematical model. Pre-RCs, composed of the origins loaded with Mcm2-7 proteins, the presumptive helicase, are formed quickly after the beginning of the origin licensing. Upon completion of the licensing phase, pre-ICs that additionally contain the essential Cdc45 are formed, first slowly and then faster, after S-Cdk has reached full activity. The assembly of complete RCs and the initiation of DNA replication from the replication origins starts around 38 minutes after the beginning of origin licensing and lasts for approximately ten minutes, denoting a very synchronous initiation of replication origins (Figure 3.10(a), lower panel). The values of the functional systems properties obtained with the reference parameter set under normal conditions are summarized in Figure 3.10(b).

3.3.2 Model simulations of characteristic S-Cdk perturbations

Four different types of perturbations of the S-Cdk activity are considered in the following: (1) a reduced concentration of S-Cdk in the nucleus, (2) a very slow degradation of Sic1 and consequently a retarded activation of S-Cdk, (3) the ΔSic1 mutant, in which S-Cdk is activated prematurely and at a reduced concentration, and (4) a stable inhibition of S-Cdk activity through

(a)



(b)

Functionality	Value
Fired origins, N	184.9 origins
Rereplicating o., ρ	0.00228 origins
Mean firing time, τ	42.40 minutes
Firing duration, Δ	4.99 minutes

Figure 3.10: (a) Simulation of the mathematical model with the reference parameter set under normal conditions. The activation kinetics of Cdks (upper panel) are identical in all generated parameter sets, since the activation of G1-Cdk is simulated by an input function and the kinetic rate constants of the reactions leading to the activation of S-Cdk are fixed. The formation kinetics of pre-RCs is also fixed, while those of pre-ICs and RCs differs in the generated parameter sets. (b) Values of the functional systems properties calculated with the reference parameter set.

a non-degradable Sic1 protein (Figure 3.11(a)-(d), upper panel). The different perturbations of the S-Cdk activity are simulated using the mathematical model with the reference parameter set (Figure 3.11(a)-(d), middle panel) and with all optimized parameter sets (Figure 3.11(a)-(d), lower panel). The functional system properties generated by reference parameter set are listed in Table 3.1 and the mean values, standard deviation and median values of all optimized parameter sets are presented in Table 3.2.

Reduced number of S-Cdk molecules

The influence of the number of S-Cdk molecules on the functional systems properties was already analyzed (Figure 3.8). Simulations with a concentration of S-Cdk reduced to one third, 200 S-Cdk molecules per cell (Figure 3.11(a)), reveal that active S-Cdk is released from inhibition by Sic1 almost 20 minutes later than under normal conditions. This results in a retarded formation of pre-ICs and a completion of RCs that is more than 20 minutes delayed. The increased time delay between origin licensing and firing leads to a very low probability for rereplication, ρ, and at the same time a high number of activated origins, N. However, the mean time to firing, τ, is almost doubled compared to normal conditions and the duration of firing, Δ, increases considerably. The extremely delayed firing of replication origins can be identified in the firing rates of all optimized parameter sets.

Table 3.1: Functional systems properties calculated with the reference parameter set for characteristic perturbations of S-Cdk activity. The corresponding values under normal conditions are specified in Figure 3.10(b).

S-Cdk activity	**Reduced**	**Retarded**	**Premature**	**Inhibited**
Firing origins N (#)	187.5	184.9	149.2	184.2
Rereplicating origins ρ (#)	6.0×10^{-11}	3.5×10^{-4}	3.7	36.6
Mean time to firing τ (min)	73.3	52.5	15.5	34.9
Firing duration Δ (min)	8.7	8.0	7.6	53.2

Slowed release of active S-Cdk

The release of active S-Cdk from inhibition by Sic1 can be slowed down, when the Cdc4-mediated recognition or the ubiquitin-mediated degradation of phosphorylated Sic1 is impaired at some step [58]. In the mathematical model, these complex processes that occur just before the release of active S-Cdk are summarized in the degradation rate of Sic1, δ_{13}. A slowed release of active S-Cdk is simulated by a 100-fold reduction of the degradation rate constant of Sic1 to 0.01 per second (Figure 3.11(b)). In this case, active S-Cdk is still released after approximately 30 minutes as under normal conditions, but at a slower rate, such that the full activity of S-Cdk is only reached after more than three hours. Simulations of this situation reveal that the activation of origins is slightly delayed and the duration of firing, Δ, is prolonged for a few minutes. Due to the small extension of the time delay between origin licensing and firing, the probability for rereplication, ρ, is decreased slightly compared to normal conditions. The same kinetic characteristics of a temporally slightly shifted and only marginally desynchronized activation of replication origins can be observed in firing rates of all optimized parameter sets.

Premature activity of S-Cdk

The inhibition of S-Cdk activity during the licensing phase is essential for an accurate duplication of the genome and the prevention of genomic instability [45]. In ΔSic1 cells, lacking the stoichiometric S-Cdk inhibitor, Sic1, S-Cdk is activated prematurely. Moreover, the concentration of S-Cdk in the nucleus is assumed to be lower than under normal conditions, due to a reduced import rate of S-Cdk into the nucleus, which is supposed to also be mediated by the Sic1 protein [69]. In the mathematical model, this situation is realized by changing the initial number of Sic1 molecules to zero and reducing the initial number of S-Cdk molecules in the nucleus to 300, simultaneously (Figure 3.11(c)). Simulations of the ΔSic1 mutant reveal that a premature activity of S-Cdk results in an activation of replication origins that starts already ten minutes after beginning of the licensing phase and even before the activation of G1-Cdk. The duration of origin firing, Δ, remains reasonably short, but the fraction of firing origins reduces strongly to only 78.5%. This is probably due to the advanced inhibition of origin (re)licensing through the phosphorylation of Orc subunits by prematurely active S-Cdk. Another fraction of the origins

(1.9%) manages to fire twice and initiates DNA rereplication, which is probably another consequence of the advanced activity of S-Cdk that also leads to an accelerated phosphorylation of Sld2 and Sld3 and a rapid formation of RCs at the origins. This opens the possibility for some origins not then inhibited by phosphorylated Orc subunits to become relicensed, since other licensing inhibiting mechanisms, mediated by G1-Cdk, are also still not operative. The strongly accelerated initiation of origin firing is a characteristic feature in all firing rates calculated with optimized parameter sets.

In experimental studies by Lengronne & Schwob (2002) [45], the ΔSic1 mutant phenotype is characterized by a reduced number of firing origins, an extended S phase, chromosomal rearrangements, and ensuing double-strand breaks. All these properties together provide a strong evidence for the emergence of genomic instability in ΔSic1 mutant cells. The mathematical model rationalizes these observations and suggests two reasons, the strongly reduced number of origins, from which the DNA synthesis is initiated, and the emergence of DNA rereplication from a few origins.

Stable inhibition of S-Cdk activity

When the stoichiometric inhibitor of S-Cdk, Sic1, is stabilized, the concentration of active S-Cdk remains very low throughout G1 and S phase. This situation was experimentally realized by a non-degradable construct of Sic1, Sic1ΔNT [89, 81]. At the same time, a preformed construct of the 11-3-2 activator complex was introduced in the cells and the ability to replicate was tested and the amount of rereplication was measured. In the mathematical model, this situation was mirrored by setting the degradation rate constant of Sic1, δ_{13}, to zero and by changing the initial number of 11-3-2 complexes, C_3, in the mathematical model from zero to 200 copies (Figure 3.11(d)). The simulations reveal that a only very small free concentration of active S-Cdk of less than 30 molecules, corresponding to approximately 5% of the total S-Cdk concentration, is present in the nuclei of these cells. However, this small amount of active S-Cdk is high enough to phosphorylate Mcm2-7 molecules bound to the origins, which is essential for the formation of pre-ICs. The first origin firing event occurs already a few minutes after the beginning of the licensing phase, but the activation of further origins is very desynchronized, probably due to the limited Mcm2-7 phosphorylation rate by S-Cdk. The firing of replication origins persists for almost 1.5 hours, in which a sizable fraction of origins (19.9%) is activated repeatedly. The firing of origins is only stopped, when rereplication inhibiting mechanisms, apart from the phosphorylation of Orc subunits by S-Cdk, which is far too slow in this case, are setting in and inhibit further relicensing of replication origin. The advanced initiation of DNA replication and the extremely slow and long lasting firing of replication origins is a characteristic of the firing rates of all optimized parameter sets.

In the experiments, a long lasting and extensive DNA replication period was observed and a considerable amount of DNA rereplication was detected in the cells, but not quantified more precisely.

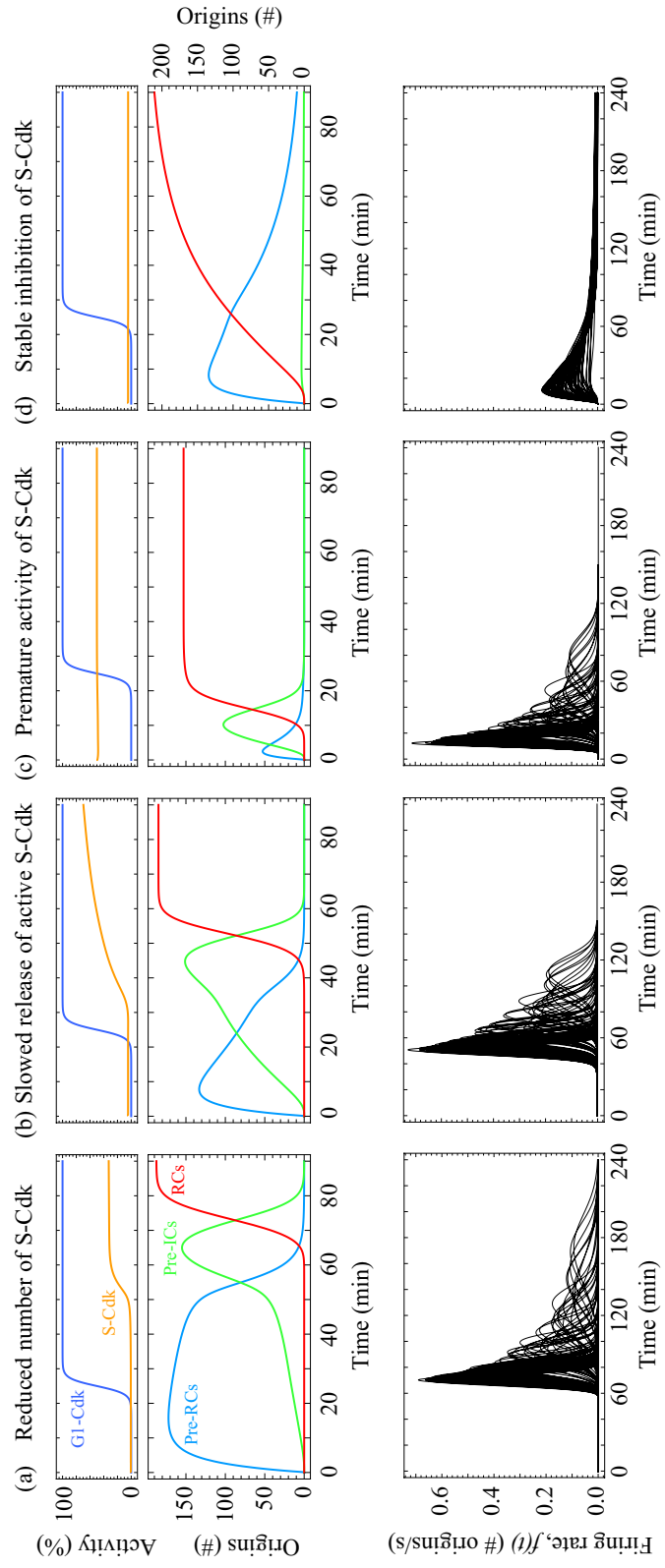


Figure 3.11: Simulations of characteristic perturbations of S-Cdk activity (upper panel). A fixed input function describes the activity of G1-Cdk (upper panel). The formation rate of pre-RCs is the same for all generated parameter sets and in all simulations, as the reaction rates in the licensing phase are fixed and independent of S-Cdk activity (middle panel). The kinetics of RC formation are simulated with the reference parameter set (middle panel). The firing rates, $f(t)$, under perturbed S-Cdk activity are derived for all generated parameter sets (lower panel).

Table 3.2: Mean values and standard deviations of the functional systems properties for different deregulations of the S-Cdk activity calculated with all optimized parameter sets (median values in brackets). The corresponding values under normal conditions are specified in Figure 3.2.

S-Cdk activity	**Reduced**	**Retarded**
Firing origins N (#)	182.8 ± 16.4 (187)	182.8 ± 1.6 (183)
Rereplicating origins ρ (#)	$(0.2 \pm 0.7\ (2.3)) \times 10^{-10}$	$(1.8 \pm 5.7\ (0.5)) \times 10^{-3}$
Mean time of firing τ (min)	86.1 ± 20.3 (76.5)	61.8 ± 14.5 (55.6)
Firing duration Δ (min)	14.7 ± 8.7 (10.3)	11.9 ± 5.5 (9.1)

S-Cdk activity	**Premature**	**Inhibited**
Firing origins N (#)	142.2 ± 7.9 (142)	174.2 ± 20.9 (181)
Rereplicating origins ρ (#)	2.0 ± 1.8 (1.8)	26.5 ± 11.8 (29.1)
Mean time of firing τ (min)	24.5 ± 15.3 (17.3)	41.8 ± 12.3 (37.5)
Firing duration Δ (min)	12.5 ± 7.4 (8.7)	58.8 ± 13.9 (55.1)

4 Analysis of the DNA replication dynamics during S phase

In this chapter, the dynamics of DNA replication in S phase is simulated by using the distribution of firing times of the early origins, $f(t)$. Quantitative properties characterizing the DNA replication dynamics in S phase, such as the duration of DNA synthesis, the distribution of DNA replicon sizes, the number of silent and the number of dormant origins, are defined and their dependencies are analyzed.

4.1 Dynamics of DNA replication in the budding yeast genome

In S phase, the whole budding yeast genome is duplicated in a semiconservative way. DNA synthesis is initiated upon completion of the RC at the replication origin. Two replication machineries, including the presumptive DNA helicases, Mcm2-7, and the DNA polymerases, are progressing bidirectionally in 3' and 5' direction from each firing origin. The DNA is copied continuously by DNA polymerase ε in the leading strand (in 3' direction), while, in the lagging strand (in 5' direction), the synthesis of new DNA proceeds discontinuously by DNA polymerase δ via the synthesis of a series of small DNA fragments that are around 100 to 200 nucleotides in length. These so-called Okazaki fragments are joined by a DNA ligase afterwards. The DNA polymerase is supposed to stop synthesis when encountering another oncoming DNA polymerase [1, 90]. The total piece of DNA that is synthesized by the two DNA polymerases belonging to replication machineries that originated from the same origin is called a DNA replicon [56]. The total amount of DNA that is replicated from a single origin is given by the size of the DNA replicon.

In every cell cycle, a fraction of origins, called silent origins, is not licensed and consequently not fired. This results in a longer piece of DNA that is synthesized by DNA polymerase starting from the neighboring origins. Another fraction of replication origins that is still licensed, but remains inactive during S phase, because of its late firing time compared to its neighboring origins is called dormant origins. Dormant origins are replicated passively by the replication machinery originating from a neighboring origin.

The duplication of the genome proceeds in a temporally coordinated manner. The major part of the DNA is replicated within a short time early in S phase from the early replication origins. If the 190 early replication origins are assumed to be distributed on the DNA according to the measured mean inter-origin distances of 46 kb [46], the total length of DNA that is replicated from these origins is $\sim$8.7 Mb, corresponding to a fraction of approximately 70% of the 12.1 Mb budding yeast genome [15].

The remaining part of the nuclear genome is synthesized from origins, that only become active late in S phase. The reasons for the delayed initiation of this part of the origins are unclear. The temporal separation of origin firing is reflected in the replication times of chromosome centers, centromeres, and chromosome ends, telomeres, which are typically replicated very early and very late in S phase, respectively (Figure 3.1(c)) [67]. The telomeric regions of the chromosomes are usually located in a dense chromatin environment [24] at the periphery of the cell nucleus [35], which might suggest a relationship between the activation time of an origin and its accessibility given by the surrounding chromatin state and its nuclear localization.

The model for the DNA replication dynamics in S phase, presented here, focusses on the early replicating part of the genome that is initiated from the early origins, whose firing kinetics is simulated with the mathematical model from Chapters 2. Possible interferences of the replication dynamics with the chromatin environment are neglected in this model. Furthermore, the progression rate of the replication fork is assumed to be equal in 3' and 5' direction and also constant, disregarding potential blockages and occasional pausing of the replication fork, due to e.g. unrepaired DNA damages.

4.2 Quantification of the replication dynamics

4.2.1 Calculation of DNA replicon sizes

To evaluate the replication dynamics during S phase, the sizes of the DNA replicons originating from all firing origins are calculated. For this, a computational model is constructed that considers a number of 190 early replication origins. The origins are positioned on the DNA strand according to the distribution of inter-origin distances that was measured by Lengronne et al. (2001) [46]. In this study, the technique of DNA combing in combination with the incorporation of exogenous bromodeoxyuridine (BrdU) into newly synthesized DNA was used to monitor the positions of the early replicating origins that became active in the presence of HU. in the computational model, the activation times of the early origins are taken from the firing rate, $f(t)$, which was calculated with the kinetic model for the initiation of DNA replication presented in Chapter 2. DNA replicon sizes are calculated for all firing origins assuming a constant, bidirectional DNA synthesis rate of 2.9 kb/min, which was determined experimentally [67]. The computational model accounts for the fraction of silent origins, which results from the action of licensing inhibitory mechanism in the kinetic model. The fraction of dormant origins is a result of the DNA replication dynamics during S phase (Figure 4.1). The calculation of the DNA replicon sizes proceeds in the following steps:

1. The positions of the 190 early replication origins, $x_0(i)$, are determined by randomly drawing distances between the adjacent origins from the measured distribution of inter-origin-distances [46].

2. An integer number of silent origins is determined stochastically according to the fraction of silent origins that was calculated by the kinetic model for the initiation of DNA replication in Chapter 2. The silent replication origins are randomly selected and removed from

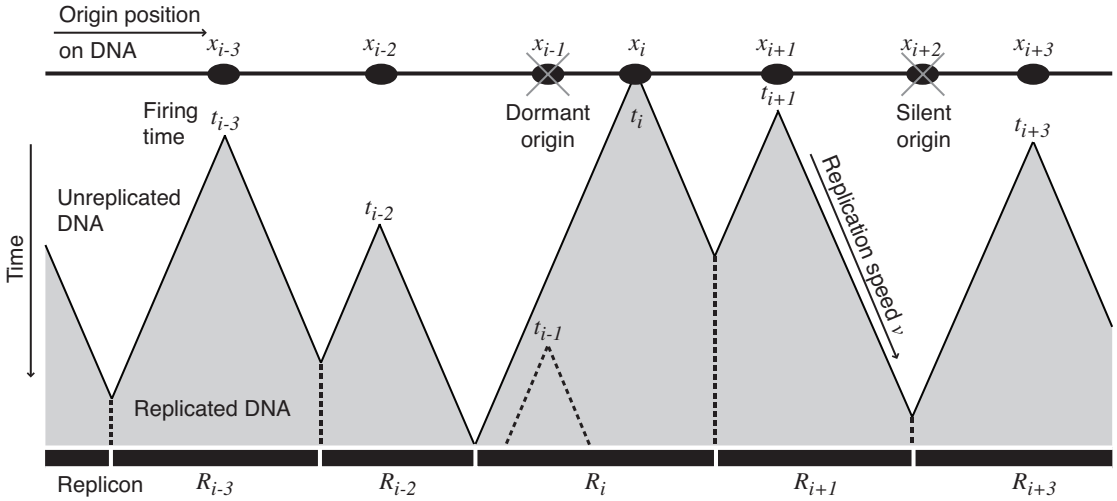


Figure 4.1: Illustration of the DNA replication dynamics during S phase. Each replicon, $R(i)$, of origin i is dependent on the origin position, x_i, the firing time, t_i, the replication speed, v, and position and firing time of its neighboring origins. Silent and dormant remain inactive during S phase for distinct reasons.

the sequence of the 190 early replication origins. This increases the distance between the two origins adjacent to a selected, silent origin. The total number of silent origins was required to be smaller than 5 % ($190 - N$ origins) in the kinetic model.

3. Each remaining origin i at position $x_0(i)$ is assigned a firing time $t_0(i)$, which is drawn randomly from the distribution of firing times of the early origins, $f(t)$. DNA synthesis is assumed to initiate simultaneously in 3' and 5' direction at the firing time $t_0(i)$.

4. By comparing the differences in the firing times, $\Delta t_0(i) = t_0(i) - t_0(i-1)$, and the time needed to synthesize the piece of DNA, $\Delta x_0(i)/v = (x_0(i) - x_0(i-1))/v$, between a pair of adjacent replication origins, i and $i-1$, with the reported DNA synthesis rate, $v = 2.9kb/min$, [67] dormant origins that are replicated passively from a neighboring replication machinery, are identified as follows:

$$\begin{aligned}
&\text{if} \quad v|\Delta t_0(i)| < \Delta x_0(i) && \text{no origin of both is replicated passively,}\\
&\text{if} \quad v|\Delta t_0(i)| \geq \Delta x_0(i) \wedge \Delta t_0(i) > 0 && \text{origin } i \text{ is replicated passively,}\\
&\text{if} \quad v|\Delta t_0(i)| \geq \Delta x_0(i) \wedge \Delta t_0(i) < 0 && \text{origin } i-1 \text{ is replicated passively.}
\end{aligned}$$

The dormant origins are also removed from the remaining sequence of early origins and the distance between the two origins adjacent to the passive origin is increased accordingly.

The identification of passively replicated origins (step 4) is repeated, until no further dormant origin is detectable.

5. For all origin i that finally become active, the sizes of the DNA replicons $R(i)$ are calculated as the sum of the DNA pieces replicated in 3' and 5' direction from that origin, $R_3(i)$ and $R_5(i)$, respectively:

$$\begin{aligned} R(i) &= R_3(i) + R_5(i) \\ &= \frac{1}{2}(\Delta x_0(i) + v\Delta t_0(i)) + \frac{1}{2}(\Delta x_0(i+1) - v\Delta t_0(i+1)) \\ &= \frac{1}{2}((x_0(i+1) - x_0(i-1)) + v(t_0(i) - t_0(i+1) - t_0(i-1))) \end{aligned}$$

4.2.2 Properties of DNA replication dynamics

To evaluate the DNA replication dynamics during S phase, characteristic properties are defined that quantify the DNA synthesis period. The model for the DNA replication dynamics, which delivers the distribution of DNA replicon sizes, is strongly dependent on the random choice of distances between neighboring origins and the random assignment of firing times to these origins. Thus, the quantitative properties of the DNA replication dynamics during S phase are determined as averages over 1000 calculations, in order to eliminate stochastic variations in the distribution of DNA replicon sizes. This average can also be interpreted as an average over a population of 1000 cells with slightly different origin activation times in each cell.

Four quantitative properties are defined to characterize the DNA replication dynamics of the early origins during S phase in budding yeast:

- The *number of silent origins*
  comprises all origins that remain unlicensed in one round of replication. The number of silent origins is determined in step 2 of the calculation of DNA replicon sizes and is specified as the mean value and standard deviation over 1000 simulations of the replication dynamics.

- The *number of dormant origins*
  comprises all origins that are still licensed, but then replicated passively from a neighboring DNA polymerase, before becoming active themselves (Figure 4.1). The number of dormant origins is determined in step 4 of the calculation of DNA replicon sizes and is specified as the mean value and standard deviation over 1000 simulations of the replication dynamics. A low number of dormant origins implies that the DNA synthesis originates from many replication origins in parallel, which contributes to the shortening of the time spent on DNA duplication.

- The *duration of the S phase*
  is the time needed for the DNA synthesis from the early origins. It is measured as the time period between the first origin firing event and the completion time of the last DNA replicon. The duration of S phase is specified as the mean value and the standard deviation over 1000 simulations of the replication dynamics. A short duration of the S phase is favorable, as it contributes to a fast progression through the cell cycle leading to a rapid cell proliferation [90].
- The *distribution of DNA replicon sizes*
  gives the fraction of DNA that is replicated from every firing origin. The distribution of DNA replicon sizes is specified as the average replicon size and its standard deviation in 1000 calculations of complete sets of DNA replicon sizes. The DNA replicons of the different origins should ideally have similar sizes indicating an equal partitioning of the DNA synthesis among the available replication start sites and the absence of very long replicon sizes that are supposed to increase the risk for dissociation or stalling of the replication fork [73, 44, 90].

The fraction of finally activated origins in one round of replication, can be interpreted as the overall origin efficiency,

$$\phi = 1 - \frac{(\#\text{ silent origins} + \#\text{ dormant origins})}{190},$$

in that S phase. The highest overall origin efficiency of $\phi = 100\%$, is achieved, when DNA replication is initiated from all origins.

4.3 Control of the DNA replication dynamics

The early S phase is the time from the activation of the first origin to the completion of the last DNA replicon replicated from an early origin. The early S phase is characterized by the distribution of DNA replicon sizes, the average duration of the early S phase and the average number of silent and dormant origins per round of replication. These S phase properties are dependent on the functional systems properties of the kinetics of origin firing that are, in parts, already determined before the activation of the replication origins. Previous analysis has shown, that the number of origins that are not licensed, $(190 - N)$, and subsequently remain silent in S phase becomes already fixed during the licensing phase, while the coherence of origin activation, Δ, is decided mostly by the kinetics of the formation of the 11-3-2 activator during the firing phase of the replication initiation network (Figure 2.1). Certainly, the DNA replication dynamics during S phase depends on the DNA synthesis rate, v. Variations in this rate may be caused by fluctuations in the availability of nucleotides and the occurrence of unrepaired DNA damages or other barriers on the unreplicated DNA strand, which lead to a pausing and stalling of the replication fork. For simplicity, the same constant progression speed of the replication fork, v, from every firing origin was assumed in the model. Measurements of the progression speed of

the replication fork show some differences, e.g. a mean values of 2.9 kb/min and 2.8 kb/min were reported by Raghuraman et al. (2001) [67] and Yabuki et al. (2002) [88], respectively, while Lengronne et al. (2001) [46] referred to a value of 3.7 kb/min.

In the following section, the dependency of the DNA replication dynamics on the number of silent origins, $(190-N)$, the coherence of origin firing, Δ, and the DNA synthesis rate, v, is analyzed. The distribution of firing times, given by the firing rate, $f(t)$, is calculated with the kinetic model for the initiation of DNA replication of Chapter 2 using the reference parameter set (Table 2.3 and 2.5) with certain modifications that are specified below.

4.3.1 Dependency on the number of activated origins N

The influence of the number of activated origins, N, on the DNA replication dynamics during the early S phase is analyzed by varying the number of activated origins, N, without changing the values of the other functional systems properties. This is realized by a strengthening of relicensing inhibiting mechanisms, that are the degradation of Cdc6, the export of Mcm2-7 proteins or the phosphorylation of Orc subunits by S-Cdk. Here, the latter mechanism is chosen to modified. Several model simulations are performed, each with a different binding rate constant of S-Cdk to Orc proteins ($7.3\times10^3(\mathrm{Ms})^{-1} < k_{26} < 1.2\times10^6(\mathrm{Ms})^{-1}$). This leads to a variation in the inhibition of origin licensing and a changed number of activated origins, N, while the other functional system properties, in particular the duration of origin firing, Δ, are not affected (Figure 4.2(b)).

The number of activated origins, N, has a strong influence on the properties of the early S phase (Figure 4.2(a)). A bisection of the number of activated origins, N, from 185 to approximately 95 origins would almost double the mean DNA replicon size to 90 kb and result in a standard deviation of the length distribution of DNA replicons of up to 140 kb. The duration of the early S phase increases enormously from around 28 minutes, under normal conditions, to approximately 70 minutes with only 95 activated origins. The number of silent origins increases according to the reduction of the number of activated origins, while the number of dormant origins is hardly influenced. However, a small reduction of the number of activated origins, N, to around 160 origins would only result in a minor deterioration of the properties of the early S phase.

This analysis reveals, that the number of firing origins, N, which is fixed already during the licensing phase, becomes very important during the S phase of the cell cycle, when the genome has to be duplicated, preferably within a short time period.

4.3.2 Dependency on the duration of origin firing Δ

Different durations of origin firing, Δ, are obtained by varying the binding rate constant of the 11-3-2 activator complex to the origins on the interval $3.8\times10^4\ (\mathrm{Ms})^{-1} < k_{23} < 9.3\times10^5\ (\mathrm{Ms})^{-1}$ in the reference parameter set and simulating the kinetic model for the initiation of origin firing. A slower binding rate constant of the 11-3-2 activator to pre-ICs at the origins results in a prolonged duration of origin activation, Δ, while the other functional system properties of the kinetics of

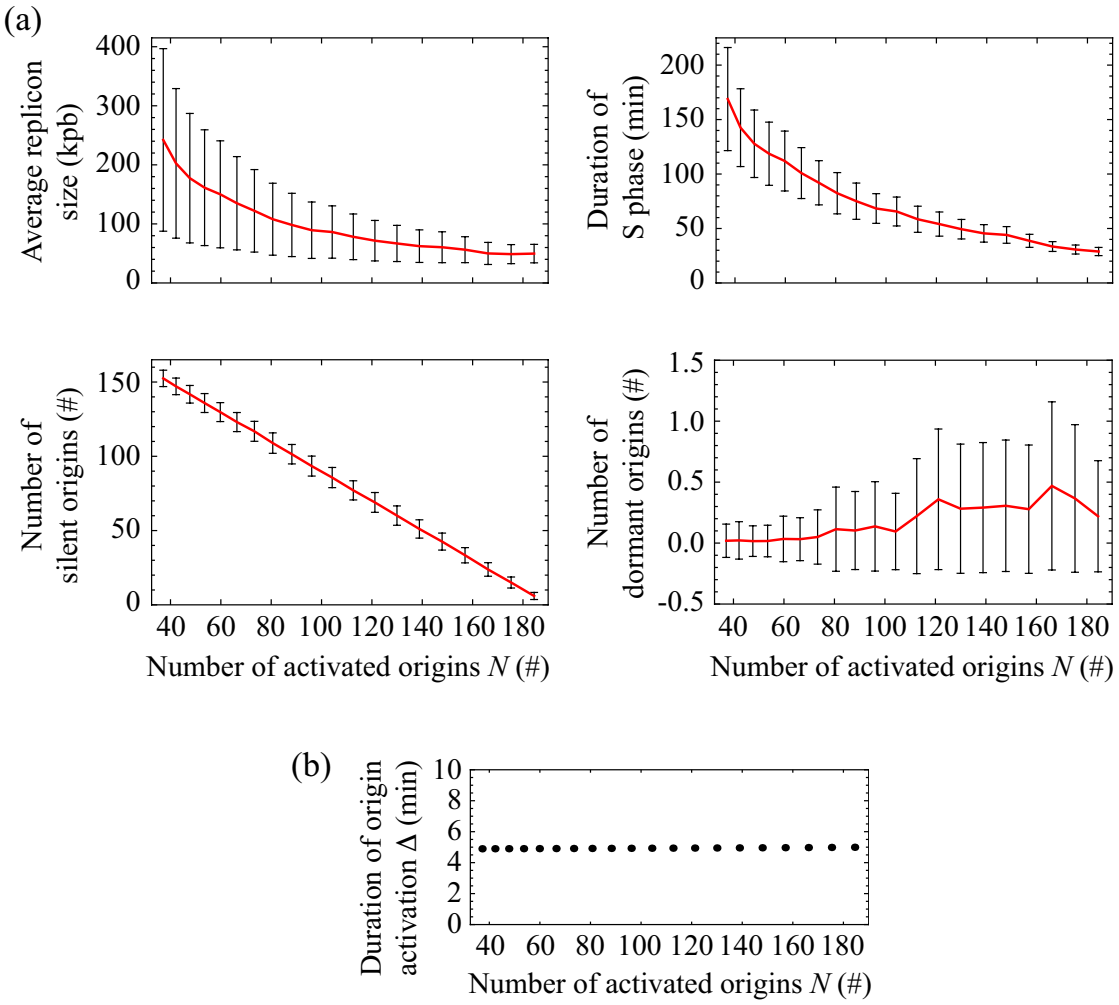


Figure 4.2: (a) Dependence of the replicon size distribution, the duration of the early S phase, the number of silent and the number of passively replicated origins on the number of activated origins, N. All S phase properties are specified by the mean value and the standard deviation over 1000 simulations. (b) The duration of firing, Δ, is unchanged by the variation in the number of activated origins, N.

origin firing are unaffected. In particular, the number of activate origins, N, remains constant 185 (Figure 4.3(b)).

A doubling of the duration of firing, Δ, (from 5 to 10 minutes) only has a small influence on the properties of the early S phase. The distribution of DNA replicon sizes would almost be unchanged, the average duration of the early S phase would increase slightly (about 10 minutes) and the number of dormant origins would rise to 10 origins per cell cycle on average. A further increase in the duration of firing, Δ, to more than 30 minutes has a strong effect on the properties of the early S phase (Figure 4.3(a)). The mean replicon size would be almost doubled (approximately 90 kb) and the standard deviation of the length distribution of DNA replicon sizes reaches values of up to 150 kb. The duration of the early S phase would be extended enormously to nearly 2 hours and the number of dormant origins increases to more than 80 origins per cell cycle. However, the number of silent origins is not affected by this variation of the duration of origin activation, Δ, and remains constant at around 5.5 origins per cell cycle.

Interestingly, the duration of the early S phase and the duration of origin firing, Δ, exhibit a nearly linear relationship, with proportionality constant of approximately three, implying that a prolongation of the duration of origin firing, Δ, of one minute causes an extension of the length of the early S phase of about three minutes.

4.3.3 Dependency on the DNA synthesis rate v

Different values for the progression rate of the replication fork, v, are reported and probably this rate varies among cells, cell cycles and even among the different replication forks. Raghuraman et al. (2001) [67] monitored the replication dynamics of the entire yeast genome and determined the average progression rate of each replication fork. Progression rates of approximately 480 replication forks were obtained, whose values are distributed between 0 and 11 kb/min with a mean value at 2.9 kb/min and a median value at 2.3 kb/min. Lengronne et al. (2001) [46] referred to a progression rate of the replication fork of 3.7 kb/min.

In the model, the progression rate of the replication fork, v, is assumed to be constant during the early S phase and to attain the same value for all replication forks. In the model, a mean progression rate of replication fork, $v = 2.9$ kb/min measured by Raghuraman et al. (2001) [67] is used in the calculation of DNA replicon sizes, the duration of early S phase and the number of silent and dormant origins. The effect of changes in the replication fork progression rate, v, on the properties of the early S phase is analyzed systematically on the range 0.5 kb/min $\leq v \leq$ 10 kb/min (Figure 4.4).

A minimal duration of the early S phase (below 20 minutes) is obtained at a very high DNA synthesis rate ($v = 10$ kb/min), but the mean replicon size and the standard deviation of the DNA replicon size distribution (60 ± 30 kb) would be increased at this high progression rate, v. The number of dormant origins is raised as well to more than 30 origins, but the number of silent origins is not influenced by a change in the progression rate of the replication fork, v, as expected.

A very tiny distribution of replicon sizes around the mean inter-origin-distance of 46 kb without the occurrence of dormant origins is obtained at a very slow DNA synthesis rate ($v = 0.5$ kb/min), but the time needed to replicate the fraction of the genome from the early origins will

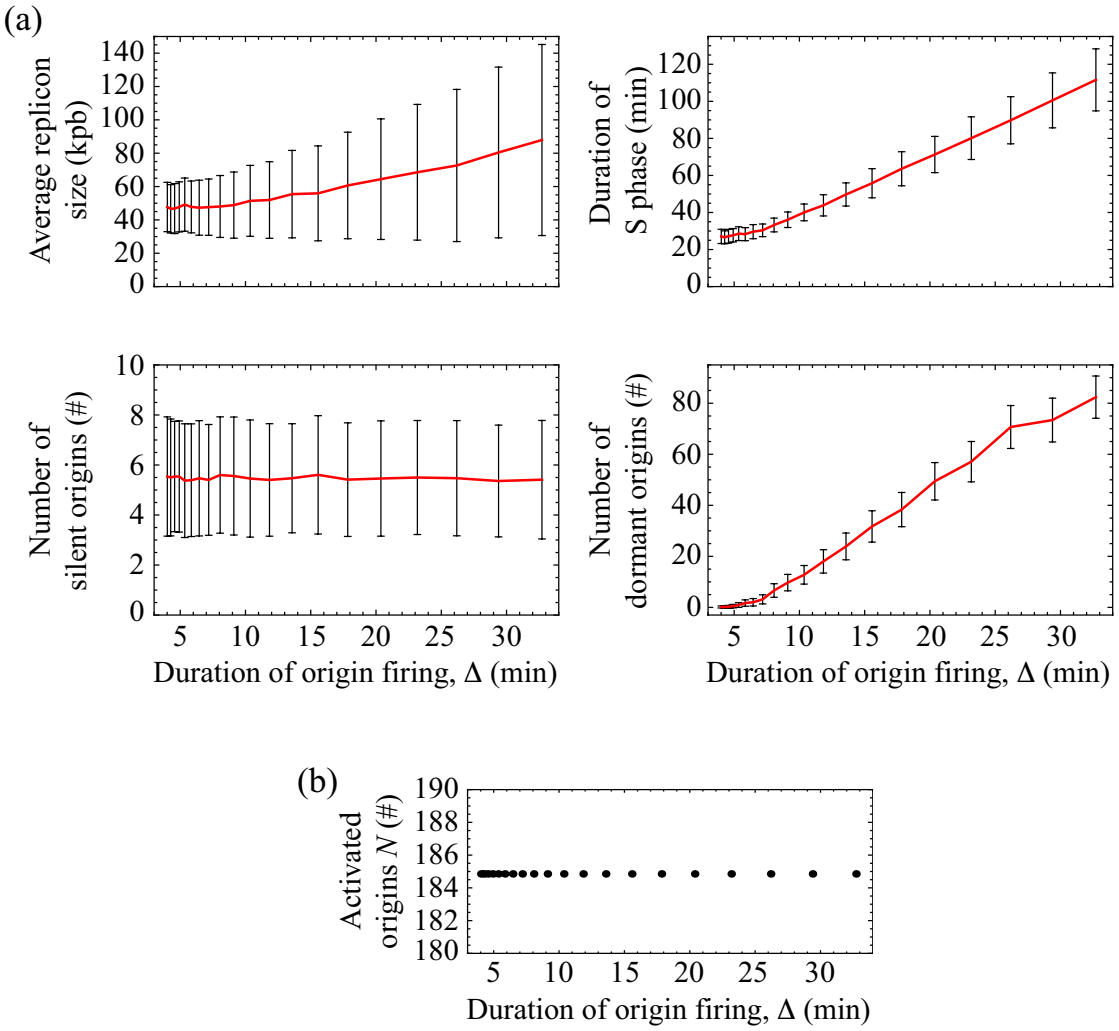


Figure 4.3: (a) Dependence of the replicon size distribution, the duration of the early S phase, the number of silent and the number of passively replicated origins on the duration of origin activation, Δ. All S phase properties are indicated by the mean value and the standard deviation over 1000 simulations. (b) The number of activated origins, N, is unchanged by the variation in the duration of firing, Δ.

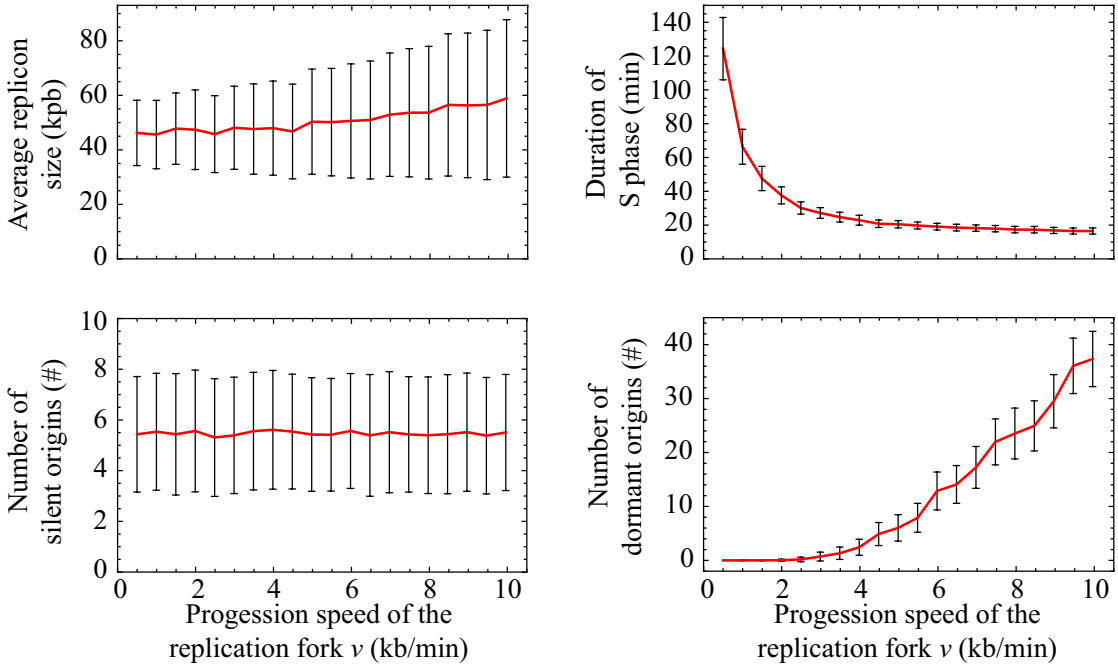


Figure 4.4: Dependence of the replicon size distribution, the duration of the early S phase, the number of silent and the number of passively replicated origins on the progression speed of the replication fork v. The distribution of origin firing events are calculated with the reference parameter set with a duration of origin activation of $\Delta = 4.99$ minutes and a number of firing origins of $N = 184.9$. All S phase properties are denoted as mean values with standard deviations over 1000 simulations of the model for the replication dynamics during S phase.

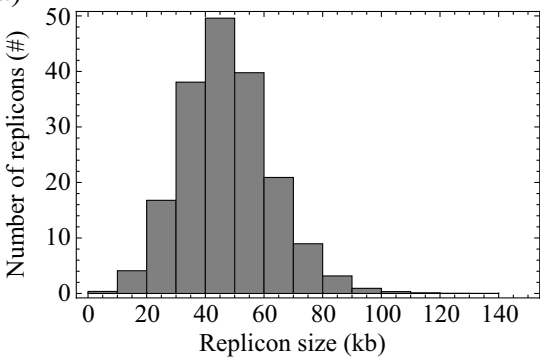
increase to more than 2 hours in this case.
In the range of the reported progression rates of the replication forks (roughly between 2 to 4 kb/min), the properties of the DNA replication dynamics during the early S phase are practically not affected. The biggest variation is observed in the duration of the early S phase, which is almost 40 minutes at a fork rate of 2 kb/min, while only 25 minutes at 4 kb/min.

4.4 Simulations of the DNA replication dynamics

4.4.1 Replication dynamics under normal conditions

The kinetics of origin firing under normal conditions was simulated in Chapter 3 for the reference parameter set. The model for the DNA replication dynamics in S phase predicts a narrow distribution of DNA replicon sizes under normal conditions. Most replicon sizes center around the mean inter-origin distance of 46 kb and the probability for the occurrence of replicon sizes

(a)



(b)

Quantitation	Value
Silent origins	5.45 ± 2.34 origins
Dormant origins	0.52 ± 0.72 origins
S phase	26.8 ± 3.35 minutes
Replicon sizes	45.60 ± 14.47 kb

Figure 4.5: (a) Distribution of DNA replicon sizes under normal conditions is calculated as an average over 1000 simulations with the firing rate, $f(t)$, obtained using the reference parameter sets. (b) Quantitative properties of the DNA replication dynamics during S phase under normal conditions. Values are again calculated as an average over 1000 simulations.

longer than 100 kb is very low (Figure 4.5(a)). The average duration of the early S phase is predicted to take around 27 minutes under normal conditions, which is consistent with the reported durations of S phases in mother and daughter cells, that are in the range between 24 and 32 minutes [11]. The mean number of dormant origins per round of replication is below zero (Figure 4.5(b)), and among 1000 cells, passive replication is detected from maximal four origins in a round replication and does not occur in the majority of cells. This implies that the multitude of origins is exploited very efficiently with an overall origin efficiency of $\phi = 97\%$ under normal conditions resulting in a DNA synthesis from many origins in parallel.

4.4.2 Replication dynamics under characteristic S-Cdk deregulations

Simulations of four different perturbations of the S-Cdk activation kinetics (Figure 3.11) were performed in Chapter 3 using the kinetic model for the initiation of DNA replication with the reference parameter set. The consequences of a deregulated S-Cdk activity on the DNA replication dynamics in S phase are quantified through the determination of the characteristic properties of the DNA synthesis period, that are, the number of silent origins, the number of dormant origins, the duration of the early S phase and the distribution of DNA replicon sizes (Figure 4.6 and Table 4.1).

Reduced number of S-Cdk molecules

A reduction of the total S-Cdk activity results in a distribution of replicon sizes that is centered narrowly around the average inter-origin distance of 46 kb and DNA replicon sizes larger than

120 kb practically do not occur. Compared to normal conditions, the number of silent origins is halved, but the fraction of dormant origins becomes substantial ($\sim 4\%$). Also the duration of the early S phase increases considerably about 8 minutes compared to normal conditions. The overall origin efficiency in cells with a reduced S-Cdk activity is only marginally impaired with $\phi = 95\%$.

Slowed release of active S-Cdk

A slowed release of active S-Cdk from inhibition by Sic1 causes similar strong changes in the DNA replication dynamics during S phase as in the case of a reduced S-Cdk activity. The overall origin efficiency is almost the same, with $\phi = 94\%$, than in cells with a reduced S-Cdk activity, but here, the number of silent origins remains as high as under normal conditions, whereas the number of dormant origins increases about 10 times compared to normal conditions. The resulting duration of the early S phase and the distribution of DNA replicon sizes is also very similar to the situation of a reduced S-Cdk activity.

Premature activity of S-Cdk

A premature activity of S-Cdk has enormous consequences on the DNA replication dynamics in S phase, which was also observed experimentally in the ΔSic1 mutant [45]. As predicted by the kinetic model for the initiation of DNA replication, the fraction of silent origins is increased to almost one fifth of all origins. Consequently, the number of dormant origins remains relatively low. The duration of S phase prolongs considerably about 20 minutes and the distribution of DNA replicon sizes is no longer centered around the mean inter-origin distance of 46 kb, but is shifted to around 60 kb. Very long replicon sizes of more than 150 kb may also occur. The overall origin efficiency reduces strongly to $\phi = 77\%$ in the ΔSic1 mutant.

Stable inhibition of S-Cdk activity

In the situation of a stable inhibition of S-Cdk activity by Sic1 and a preformed 11-3-2 activator complex, which was experimentally realized by Zegerman and Diffley (2007) [89], the DNA replication dynamics during S phase is entirely disordered. While the number of silent origins is not changed compared to normal conditions, the number of dormant origins is increased enormously, due to the very unsynchronized firing of replication origins. The model predicts that almost half of the origins is replicated passively and the average duration of the early S phase prolongs to more than 100 minutes. The DNA replicon size averages around 90 kb and DNA replicon sizes longer than 200 kb may occur. The overall origin efficiency of $\phi = 49\%$ implies that only every second origin would be finally activated to initiate DNA replication, in this situation.

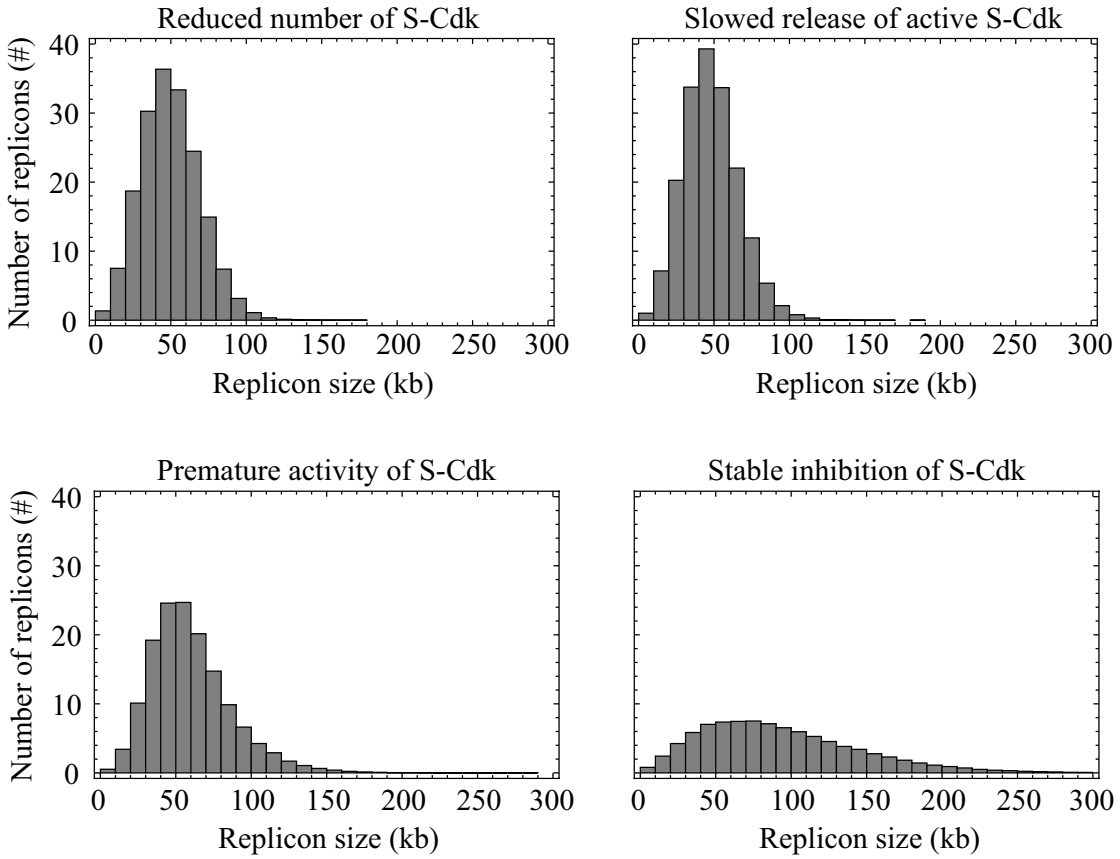


Figure 4.6: Distribution of replicon sizes caused by the origin firing kinetics calculated with the reference parameter set for different perturbations of S-Cdk activity (shown in Figure 3.11). The distribution is determined as an averaged over 1000 calculations.

4.4.3 Replication dynamics of early and late firing origins

The whole budding yeast genome is replicated in a temporally coordinated manner. A major part of the genome is duplicated from origins that become active early, while other parts are replicated from origins that become active late in S phase. The reasons for this separation in late and early origins are unknown as well as, whether and which biochemical determinants exist that delay firing of late origins. In metazoans, the duration of the S phase is prolonged with the development from embryonic to somatic cells. Concomitant with this, a decrease is observed in the number of activated origins and an increase in the asynchrony of origin firing, in which the activation of early and late origins becomes more and more separated in time [90]. During this development, a modification of the chromatin state occurs, which is accompanied with a compactification of specific regions in the chromatin [47]. Together with the fact that telomeric regions, that are also densely packed, always replicate late in S phase, a relationship is suggested between the chromatin state, that is the accessibility of an origin and its replication time [24].

The early and late replicating origins are licensed both at the same time, during the licensing

Table 4.1: Quantification of the DNA replication dynamics during S phase calculated as an averaged over 1000 simulations for characteristic S-Cdk perturbations using the reference parameter set

S-Cdk activity	**Reduced**	**Retarded**	**Premature**	**Inhibited**
Silent origins (#)	2.45 ± 1.47	5.56 ± 2.27	40.85 ± 5.53	5.52 ± 2.34
Dormant origins (#)	6.97 ± 2.44	5.07 ± 2.10	2.25 ± 1.46	90.87 ± 6.24
S phase (min)	34.76 ± 3.56	33.64 ± 3.54	45.64 ± 7.58	102.35 ± 14.17
Replicon sizes (kb)	50.15 ± 18.82	48.60 ± 18.16	61.13 ± 27.02	93.45 ± 52.57

phase, and only differ in their time of replication initiation [71]. The fact that DNA rereplication has also been observed from late firing origins [82] confirms the idea that the activation of both, early and late firing origins, is regulated by similar mechanisms, although additional factors might be involved in the activation of these origins [85].

The parameter sets that were generated through the optimization of the systems functionality in Chapter 2 all have identical licensing parameter values but varied firing parameter values that resulted in the different kinetics of origin firing (Figure 3.1). Most optimized parameter sets exhibited a fast firing kinetics, but a few generated parameter set gave rise to a more slowed initiation of origin firing.

The different firing rates, $f(t)$, generated by the different optimized parameter sets may also be interpreted as the activation kinetics of early and late origins and can be utilized to analyze the DNA replication dynamics of a complete S phase.

To simulate the firing kinetics of the late origins, a second reference parameter set is selected from the ensemble of optimizes parameter sets. In agreement with experimental observations [88], a representative parameter set is chosen with average initiation time of the late origins that is around 10 minutes delayed compared to the average activation time of the early origins. If the spacing of replication origins is assumed to be the same for the early and the late firing origins, approximately 80 late firing origins are needed to replicate the remaining part of the genome of around 3.6 Mb (granted that the whole budding yeast genome consists of 12.2 Mb and that the 190 early origins that are distributed with a mean inter-origin-distance of 46 kb [60, 46], are capable to replicate 8.7 Mb of the genome, on average). The firing rates, $f(t)$, calculated for 190 early and 80 late origins are shown in Figure 4.7.

The difference between an early and a late firing origins in terms of kinetic parameter values is not well-defined. Deviations are identifiable in almost all parameter values of the firing module, which makes the identification of a specific reaction as the reason for the differences in firing times of early and late origins impossible.

For the evaluation of the DNA replication dynamics during a complete S phase (Table 4.2), the simplifying assumption is made that the early and late firing origins are totally separated within the genome and that their activation times are independent from each other.

The duration of a total S phase is estimated to take around 40 minutes, which could still be

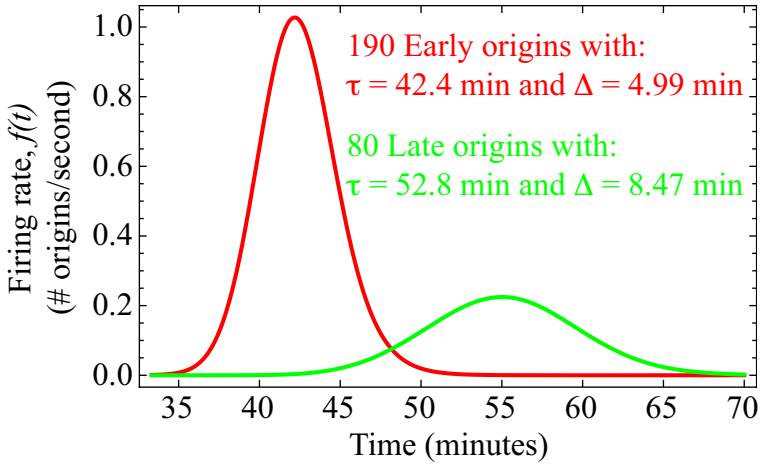


Figure 4.7: One fast and one slow parameter set that both generate a relatively synchronous firing kinetics are selected to describe the kinetics of early and late origin activation.

Table 4.2: Predicted quantitative properties of the DNA replication dynamics during a complete S phase considering both, early and late firing origins.

Quantitative property	**Early origins**	**Late origins**	**All origins**
Initial origins (#)	190	80	270
Silent origins (#)	5.36 ± 2.30	1.99 ± 1.38	7.07 ± 4.03
Dormant origins (#)	0.66 ± 0.78	4.05 ± 1.96	6.66 ± 7.26
S phase (minutes)	27.28 ± 3.61	31.56 ± 3.55	40.77 ± 3.44
Replicon sizes (kb)	46.72 ± 15.03	47.95 ± 18.94	47.07 ± 16.25

reduced, if the early and late replicating territories are assumed to be mixed, such that the early replicating part of the genome is interspersed with smaller groups of late replicating origins, as is probably the case in each of the chromosomes in budding yeast [67].

5 Discussion and Outlook

In this work, a comprehensive mathematical model is provided for the regulation of DNA replication in budding yeast, starting with the formation of RCs at the replication origins and ending after the completion of DNA synthesis. We considered all measured data reported up to now, and constructed a molecular reaction network based on the identified molecular interactions. Experimentally determined molecule concentrations, albeit in an asynchronous cell population, are used as initial values to simulate the kinetic model. Many kinetic parameters, such as binding, dissociation, (de)phosphorylation, degradation and export rate constants have not been determined experimentally.

To identify these kinetic parameter values computationally, we used an optimization approach to maximize the functionality of the system, that is the rapid and synchronous activation of replication origins without rereplication, in dependence of the kinetic parameter values that were only allowed to vary within biochemically reasonable ranges.

The systems functionality is defined by means of four functional systems properties, the number of activated origins, the number of rereplicating origins, the mean time to origin firing, and the duration of firing. These four systems properties quantify the kinetics of origin firing, which is the output of the mathematical model, in a comprehensive and biologically interpretable way. Contrariwise, the choice of limits of these systems properties that distinguish a still functional kinetics of origin firing from an already non-functional kinetics is undetermined. No quantitative values can be found for these limits in experimental data, which often only comprise qualitative observations. We have chosen reasonable limits for the functional system properties, which we used as constraints in the parameter optimization algorithm.

Analysis of the DNA replication initiation network reveals the major controlling molecules and reactions of each functional systems property. Additionally, correlations between these properties are identified. The impact of multiple phosphorylation on the initiation of DNA replication is elucidated by the mathematical model. A specific function is ascribed to the multisite phosphorylation of Sic1 and Sld2 in creating a time delay between G1-Cdk and S-Cdk activity, in the case of Sic1, and between S-Cdk activation and the activation of replication origins, in the case of Sld2. Moreover, experimentally realized deregulations in the replication initiation network are simulated and quantified using the mathematical model.

Finally, the consequences of a particular kinetics of origin firing on the DNA replication dynamics during S phase are analyzed in a second mathematical model. For a given progression rate of the replication fork, the number of activated origins and the duration of origin firing, derived from the kinetic model for the initiation of DNA replication, become very important during S phase. Both strongly influence the total time spent on the synthesis of DNA and the prevalence of passive replication.

Using these two mathematical models, one for the kinetics of DNA replication initiation and one for the DNA replication dynamics during S phase, which are both based upon certain sim-

plifying assumptions, we are able to analyze the process of DNA replication in budding yeast in a quantitative and comprehensive way.

Specific points concerning the construction of the mathematical model or the biological interpretation of the results are discussed in the following.

5.1 Construction of the mathematical model

5.1.1 Structure of the molecular interaction network

The molecular interaction network is constructed on the basis of a multitude of experimental data, however, details about certain reactions, such as their temporal order, are still discussed. In the mathematical model, some additional assumptions have been made, while other aspects have been neglected.

The mathematical model proposes and strongly requires the recycling and a catalytic action of the 11-3-2 activator complex for its functionality. Without this recycling, the number of 11-3-2 activator complexes, which is limited by the experimentally determined amount of Sld3 molecules [27], would not be sufficient for the initiation of more than 65 replication origins. The constituents of the 11-3-2 activator complex, Sld2, Dpb11 and Sld3, are reported to not be part of the replication complex [10], implying a release of the 11-3-2 activator complex, possibly after its catalytic action.

For thermodynamic reasons, a catalytic function of the 11-3-2 activator complex is required, since a catalytic modification of the GINS–DNA polymerase complex at the origins by the 11-3-2 activator complex would explain its dissociation only after the loading of GINS complex and DNA polymerase [3]. Hence, the proposed role of the 11-3-2 activator complex in the firing phase may be similar to those of Cdc6 in the loading of the putative DNA helicase, Mcm2-7, in the licensing phase [68]. Up to now, a catalytic function of the 11-3-2 activator complex has not been identified experimentally.

An alternative binding scheme to the complete formation of the 11-3-2 activator complex in the nucleoplasm is discussed, in which first Sld3 and Cdc45 bind together to the origins, and then Sld2 and Dpb11 form a complex in the nucleoplasm, which binds later, completing the 11-3-2 activator complex only at the origins [80]. Consequently, the phosphorylation of Sld2 by S-Cdk, which is required for the binding to Dpb11, occurs exclusively in the nucleoplasm, whereas Sld3 might be phosphorylated by S-Cdk either at the origins or in the nucleoplasm. As this alternative binding model still comprises a delayed completion of the 11-3-2 activator complex, albeit exclusively at the origins, it would not change the kinetics of origin firing and the resulting sensitivity coefficients.

Detailed information would be advantageous for the mathematical model, but is still missing for certain reactions. It is not known, whether Cdt1 needs to be phosphorylated at all by G1-Cdk or S-Cdk in order to be exported from the nucleus. The sequence of binding events of Sld2 and Sld3 to Dpb11, forming the 11-3-2 activator complex is discussed [80], and also the precise number and sequence of phosphorylation events in Sld2 by S-Cdk is undetermined to a certain

extent [54, 77]. In the case of absence of exact information on a reaction, we generally assumed the simplest possible reaction scheme in the mathematical model.

Some details of reactions that are already confirmed experimentally are not considered or simplified in the mathematical model. Mcm2-7 proteins, bound to the origins, are likely phosphorylated at multiple residues by Ddk [72]. As this phosphorylation of Mcm2-7 proteins by Ddk is reported to proceed in a processive way, that is, the kinase stays attached to the substrate while phosphorylating, we summarized this process in a single phosphorylation, possibly of longer duration.

Moreover, the binding of Clb5 to Orc6 after origin activation, which is supposed to also contribute to the inhibition of origin relicensing once firing is initiated [84], was not considered in the mathematical model, as this process is still discussed and details about it are still lacking. Additional regulations of Cdc6, which also contribute to the prevention of DNA rereplication, like the regulation of its transcription [36], are also neglected in the mathematical model.
Overall, the prevention of DNA rereplication is incorporated realistically in the mathematical model by the three processes that are taken into account, the degradation of Cdc6, the export of Mcm2-7 proteins, and the phosphorylation of Orc subunits by S-Cdk, which all have also been examined experimentally [59, 84, 31].

5.1.2 Definition of the systems functionality

The principle output of the mathematical model is the firing rate, $f(t)$, which gives the distribution of activation times of the replication origins. The properties of the firing rate are mathematically well defined by the zeroth, first, and second moment of the distribution of origin firing times, $\mathscr{M}_0, \mathscr{M}_1$, and $\mathscr{M}_2$. From this, the definition of four functional systems properties resulted (Table 5.1). Although these functional systems properties have a clear biological interpretation, the specification of functionality criteria, separating functional systems from non-functional systems, is undetermined. Obviously, a high total number of activated origins is beneficial, while a low number of rereplicating origins and a rapid rate of origin firing is advantageous. Since most experimental observations reported are more qualitative, we have chosen reasonable values as limits for the functional systems properties to define the functionality of the system (Table 5.1).

The choice of these limits is reasonable but in a way arbitrary. The total number of early origins was determined by Lengronne et al. (2001) [46] to 190 origins, out of 400 origins in total. If we start initially with only 190 early origins in our model, the requirement, that almost all origins ($\sim 95\%$) should be activated finally, makes sense, but if we start the model with all $\sim$400 origins, the early activation of only the early origins ($\sim 50\%$) would be appropriate.

The number of rereplicating origins was required to be less than one reinitiated origin in every 100^{th} cell cycle for a functional kinetics of origin firing. Unfortunately, there are no experimental data available indicating that this is a valid assumption. The experimental techniques are not sensitive enough to detect an amount of rereplication of less than $\sim 1\%$ by now [31, 2], which would correspond to ρ of approximately 1-2 origins. Thus, it is totally unclear, if the limit of 0.01 rereplicating origins is a good estimate, or if the number of rereplicating origins, in fact, has to be much lower (e.g. $< 10^{-5}$) in a functional initiation of DNA replication. Moreover, it is not

Table 5.1: Functional systems properties defined by the zeroth, first and seconds moment and the chosen criteria for functional systems.

Functional systems property	Definition	Criterion
Total number of firing origins	$N = \mathscr{M}_0{}^*$ *: system without rereplication	$N > 180$ origins
Number of rereplicating origins	$\rho = \mathscr{M}_0 - \mathscr{M}_0{}^*$	$\rho < 0.01$ origins
Mean time to firing	$\tau = \mathscr{M}_1/\mathscr{M}_0$	$\tau < 85$ minutes
Duration of firing	$\Delta = 2\sqrt{\mathscr{M}_2/\mathscr{M}_0 - (\mathscr{M}_1/\mathscr{M}_0)^2}$	$\Delta < 25$ minutes

know, whether and to which extent a very low amount of rereplication, below the detection limit, has negative effects on the viability of budding yeast cells or contributes to genomic instability [31, 2].

The limiting criteria for the timing of origin firing are more straight to define. The duration of an average cell cycle was determined experimentally to take between 90 and 200 minutes [21], though a time period for the initiation of DNA replication of 85 minutes at most seems reasonable for a functional system. Experimental evidence also exists for the duration of early origin firing [67, 88], which was determined to be in the range of 16 to 20 minutes, justifying the chosen maximal limit of 25 minutes.

5.1.3 Optimization of the systems functionality

The determination of functional parameter sets depends on the chosen criteria for the definition of the functionality of a system. However, the resulting findings concerning kinetic behavior and control of the DNA replication initiation system would also be valid if other values for the functionality criteria would have been taken.

The compromise between rapid origin firing and the accurate prevention of DNA rereplication is identifiable, even if the constraints for the functional systems properties would have been chosen differently in the optimization process. In particular, a lower requirement for minimal number of activated origins, N, in the parameter optimization process, does not shift the lower limit for the mean time to origin firing, τ, below 40 minutes (and the duration of origin firing, Δ, below 4 minutes), implying, that this might be a general limit for the rate of origin activation (Figure 5.1(a)).

Functional parameter sets can be generated, in which the limit for the maximum admissible number of rereplicating origins, ρ, which is not defined properly by experimental measurements, is reduced. A variation of this constraint for the number rereplicating origins, ρ, over several

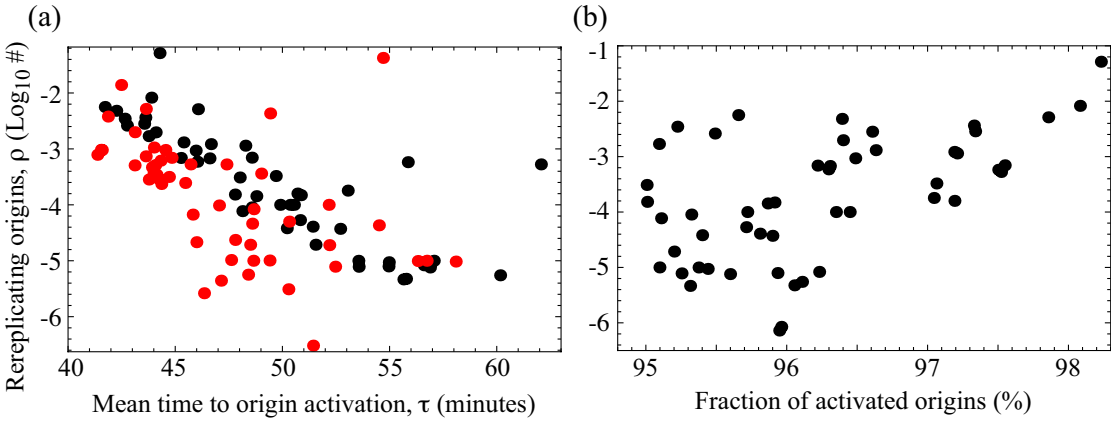


Figure 5.1: (a) Parameter sets that are generated using the reference criterion (Table 5.1) for the number of activated origins, N, and a varied limit for the maximum allowed number of rereplicating origins, $\rho < \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$, (black dots), as well as parameter sets that are generated using a relaxed constraint for the required fraction of activated origins, $N > 50\%$, with the same range of limits for the maximum allowed number of rereplicating origins, ρ, (red dots), both exhibit a correlation between the mean time to origin firing, τ, and the number of rereplicating origins, ρ, (correlation coefficient for the black dots: -0.561 and for the red dots: -0.528) confirming that faster parameter sets generally tend to exhibit a higher probability for rereplication, ρ. (b) Changing the constraint for the number of rereplicating origins, $\rho < \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$, at a fixed constraint for the fraction of activated origins, $N < 95\%$, in the parameter optimization process, implies that the maximal achievable number of activated origins, N, is limited by the maximum admissible number of rereplicating origins, ρ.

orders of magnitude in the optimization process, demonstrates that the maximum achievable number of activated origins, N, is reduced in the case of a lower limit for the probability of rereplication, ρ (Figure 5.1(b)). A lower number of firing origins, N, simultaneously implies a higher amount of origins that are actively inhibited in becoming licensed.

As the exact limits of the functional systems properties are not known and probably even vary from cell to cell, we did not intend to find a single optimal parameter set. Instead, we generated a population of more or less equally functional parameter sets. To analyze the systems behavior in response to specific perturbations, we always referred to the whole population of functional parameter sets, which would correspond to the experimental situation in which measurements are conducted in a cell population.

Analysis of the DNA replication dynamics in S phase revealed, that a short duration of origin firing, Δ, becomes very important during the DNA synthesis process, confirming that the

minimization of the duration of origin firing, Δ, as an objective function in the optimization algorithm, is a reasonable approach. Additional functionality requirements are incorporated as constraints for the values of the other functional systems properties in the optimization process. In certain cases, objective function and side constraints are exchangeable in the optimization process, resulting in equivalent parameter sets [34]. An alternative optimization approach would have been, here, to minimize the number of rereplicating origins, ρ as an objective function and to specify a limit for the duration of firing, Δ, as a constraint.

5.2 Biological insight from the mathematical model

5.2.1 Limitations and correlations of the functional systems properties

The analysis of the mathematical model for the initiation of DNA replication in budding yeast revealed correlations of the different functional systems properties. While the mean time to origin firing, τ, and the duration of origin firing, Δ, exhibit an almost linear relationship, the logarithm of the number of rereplicating origins, ρ, and the timing of origin firing, τ and Δ, appear to be anti-correlated. The latter finding is interpretable as a trade-off between the accuracy and the velocity of the initiation of DNA replication at the origins. In experiments by Zegerman & Diffley (2007) [89] and Tanaka et al. (2007) [81], in which the activation of replication origins was accelerated by the addition of a preformed 11-3-2 activator complex to the cells, a substantial amount of DNA rereplication was detected, but, unfortunately, not quantified more precisely. This observation confirms the theoretical identification of a conflict between the rapid initiation of DNA replication and the reliable inhibition of DNA rereplication.

The regulation of the inhibition of DNA rereplication emerged to be most sensitive towards fluctuations in the molecule concentrations and the kinetic rate constants. Unfortunately, this prediction can not be supported experimentally, as the total amount of DNA rereplication is still very small and in most cases below the detection limit of current experimental methods. However, it is discussed, whether additional DNA rereplication, below the up to now measurable amount, is existing [31].

Due to this restricted experimental data on the occurrence of DNA rereplication, we did not attempt to make statements about the redundancy or necessity of the different mechanisms for the inhibition of DNA rereplication. A realistic comparison of these different mechanisms is not possible, as the exact kinetic rate constants for the reactions preventing DNA rereplication are not available. In the mathematical model, we tuned these kinetic rate constants to account for experimentally examined situations, in which specific rereplication inhibiting mechanisms are affected resulting in a substantial amount of DNA rereplication [2, 37]. Certainly, all three mechanisms for the inhibition of DNA rereplication that are considered in the mathematical model, together have the ability to prevent DNA rereplication adequately.

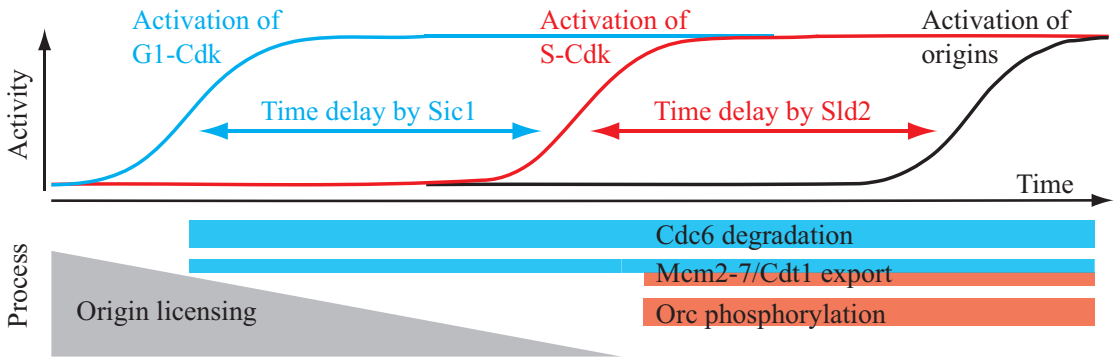


Figure 5.2: The licensing and firing of replication origins occurs at different times that are temporally robustly separated by two time delays. The first is generated by the multiple phosphorylation of Sic1 by G1-Cdk leading to the activation of S-Cdk and the second time delay is due to the multisite phosphorylation of Sld2 by S-Cdk finally resulting in the activation of the replication origins. In this way, the reactivation of an already fired origin is inhibited reliably.

5.2.2 Importance of multisite phosphorylation in creating a time delay

Analysis of the mathematical model confirmed the importance of the multiple phosphorylation of Sic1 [20, 58] and Sld2 [54, 77]. By studying in detail the influence of the number of phosphorylation sites in Sic1 and Sld2 on the kinetics of origin firing, we identified a specific role of each multisite phosphorylation (Figure 5.2). Both, multiple phosphorylation of Sic1 and Sld2, provide an essential time delay in the assembly of replication complexes at the origins. The time delay generated by Sic1 separates the activation of G1-Cdk from the activation of S-Cdk, maintaining a sufficient period of low S-Cdk activity. The multisite phosphorylation of Sld2, on the other hand, creates a time delay at the time of S-Cdk activation, which separates the activation of S-Cdk from the activation of origin firing, ensuring a sufficient time period for the execution of the G1-Cdk-dependent and, in particular, the S-Cdk-dependent mechanisms for the inhibition of DNA rereplication, before the onset of origin firing. As the S-Cdk-dependent reactions preventing DNA rereplication are exclusively executed during the second time delay, when S-Cdk is already fully active (Figure 5.2), the contribution of the multisite phosphorylation of Sld2 to the overall robustness against DNA rereplication is slightly stronger than the contribution of the time delay generated by the multisite phosphorylation of Sic1. Indeed, the importance of the multisite phosphorylation of Sic1 has been questioned recently [17, 32].

If the G1-Cdk-dependent mechanisms for the prevention of DNA rereplication are assumed to have faster kinetic rate constants, the total time delay provided by the multiple phosphorylation of Sic1 and Sld2 could be shortened without impairment of the robustness against DNA rereplication, but probably at the expense of the total number of licensed and subsequently activated origins.

The suggested role for multisite phosphorylation in the robust timing and synchronization of molecular events may also extend to other cell-cycle regulators.

5.2.3 Consequences on the DNA replication dynamics during S phase

The mathematical model for the DNA replication dynamics during S phase is also based upon certain simplifications. All early origins are assumed to be localized in a single connected cluster, neglecting the partitioning of the genome into 16 chromosomes in budding yeast and the distribution of the origins onto these chromosomes. The fork progression rates are taken to be equal for all replication forks and stalling of the replication machinery is generally not considered in the model. Experimental observations indicated that the progression rates of the individual replication forks actually vary [67]. Despite of these simplifying assumptions, the model allows capable predictions about the DNA replication dynamics during S phase and realistic estimations about its sensitivities towards specific factors.

While the impact of the replication fork progression rate and of the number of activated origins on the DNA replication dynamics is quite obvious, an effect of the duration of origin firing is more difficult to assess. A short mean time to firing, τ, and a minimized duration of origin firing, Δ, together, ensure a rapid progression through the process of genome duplication, which consists of the initiation of DNA replication at the origins and the synthesis of DNA. Thus, the duration of the DNA synthesis period is an important property of the S phase.

In the kinetic model for the initiation of DNA replication, licensing of replication origins can be inhibited through diverse mechanisms that thereby prevent DNA rereplication. There is no possibility to suppress the firing of an already licensed origin in the kinetic model, such that the number of licensed origins is always identical to the number of firing origins. The model for the DNA replication dynamics in S phase realizes a situation, in which a licensed origin finally is not activated due to a passive replication by a neighboring replication machinery. The passive replication of origins is reported to cause a transient pausing of the replication fork at the passively replicated origin [90] and a lengthening of the DNA replicons replicated by the neighboring replication machineries, but it is not clear, whether this bears an increased risk for the stalling of the replication fork. The overall origin efficiency, which is affected by both, the number of unlicensed (silent) origins and the number of passively replicated (dormant) origins, is another important property of the S phase.

Certainly, the length of the DNA replicons is also an important property of the S phase. Long DNA replicons probably exhibit several disadvantages, like an increased fragility for the accumulation of tensions in the DNA double helix and for stalling of the replication fork, which might activate certain checkpoint responses [73].

All important properties of the DNA replication dynamics in S phase are shown to be dependent on the number of activated origins, N, as well as on the duration of origin firing, Δ, which become fixed in the licensing and firing phase, respectively.

In summary, the strongest influence on the properties of the S phase is exerted by the number of activated origins, N, on the duration of DNA synthesis and on the average DNA replicon size, while the overall origin efficiency is only moderately affected (Table 5.2). A change in the duration of origin activation, Δ, mostly influences the duration of S phase and slightly effects the average replicon size and the overall origin efficiency. Generally, the number of activated origins, N, and the duration of origin firing, Δ, have opposing effects on the properties of the S

Table 5.2: Control coefficients for a 1% change in the number of activated origins, N, and the duration of origin firing, Δ.

S phase property	Activated origins (N)	Duration of origin firing (Δ)
Duration of DNA synthesis	- 2.52	1.07
Average DNA replicon size	- 2.56	0.55
Overall origin efficiency	1.42	- 0.48

phase.

5.3 Outlook

Two mathematical models are developed in this work to describe the process of genome duplication in budding yeast, from the licensing of replication origins to the completion of DNA synthesis. The models may be improved and extended in several directions.

The present mathematical models would greatly benefit from more experimental data, such as phosphorylation kinetics of network components or binding affinities. Quantitative and more sensitive measurements of the amount of DNA rereplication would render the predictions of the kinetic model even more realistic. Possibly, this would also enable statements on the redundancy of the mechanisms preventing DNA rereplication and estimations of the impact of a specific mechanism on the accurate inhibition of DNA rereplication.

The parameter optimization approach could also be applied to the kinetic parameters of the licensing and S-Cdk activation modules, to analyze the kinetics of the licensing phase in more detail. This might confirm the identified interrelations of the functional systems properties and, potentially, uncover further correlations. Possibly, this would also provide some new insight into the choice of limits of the functional systems properties, which have to specified in the optimization algorithm.

The theoretically predicted role of the multisite phosphorylation of Sld2 by S-Cdk, in providing a time delay for the separation of S-Cdk activation and origin activation could also be tested experimentally.
We would propose to introduce an Sld2 mutant that is rendered constitutively active by substituting threonine 84 with aspartic acid, or to introduce a preformed 11-3-2 activator complex, such as the construct that is used in experiments by Tanaka et al. (2007) [81] and Zegerman & Diffley (2007) [89], into budding yeast cells with a normal S-Cdk activity and activation following the degradation of Sic1. The model predicts a premature initiation of DNA replication in this situation, leading to a desynchronized activation of replication origins and some amount of DNA rereplication (from around 10 origins), although the S-Cdk-dependent mechanisms that prevent DNA rereplication are unaffected. Thus, this experiment could potentially clarify, if the observed considerable amount of DNA rereplication is due to a lack of inhibitory S-Cdk activity in the experiments by Tanaka et al. (2007) [81] and Zegerman & Diffley (2007) [89] or due to the temporally deregulated initiation of DNA replication. The model would propose the latter reason for the DNA rereplication.

The mathematical model could be extended by incorporating also the late replication origins into the model. The late firing origins are licensed simultaneously with the early firing origins, but the regulation of their firing is probably influenced by additional not yet fully understood processes (see section 4.5). A complete model, considering the activation of all replication origins, could be used to analyze potential dependencies of the activation kinetics of early firing origins on those of the late firing origins and vice versa.

The DNA replication dynamics during S phase could be improved by the partitioning of the origins on the 16 chromosomes of the budding yeast genome and by the consideration of the

corresponding lengths of each chromosome.

A direct combination of both models would allow to utilize the predicted completion times of the DNA replicons in the kinetic model for the initiation of DNA replication. In this case, the release of the molecular constituents of the RC after completed DNA synthesis, and the reentering of an already fired origin into the replication initiation network at state S_0 in the kinetic model could be taken from the model for the DNA replication dynamics during S phase.

Budding yeast is used here as a model organism, since a lot of molecular details on the regulation of the activation of replication origins have been identified experimentally in budding yeast. This facilitates the construction of a realistic and comprehensive mathematical model for the initiation of DNA replication.

The basic regulatory mechanism responsible for the synchronous and precise activation of replication origins, which has been quantitatively analyzed in this work, could similarly be present in higher eukaryotic cell types. Indeed, many proteins of the replication initiation network, such as Sld2 or the licensing factors, Cdc6 and Cdt1, are conserved from yeast to higher eukaryotes. Mathematical modeling of the initiation of DNA replication in other cell types, could be very helpful to identify these mechanisms and to characterize the initiation of DNA replication in these cell types.

Abbreviations

Table .3: Name descriptions and alias names of molecules mentioned in this work

Molecule	**Alias**	**Name Description**
Cdc6		**C**ell **D**ivision **C**ycle 6
Cdc28	Cdk1	**C**ell **D**ivision **C**ycle 28
Cdc45	Sld4	**C**ell **D**ivision **C**ycle 45
Cdt1	Tah11	**T**opo-**A** **H**ypersensitive 11
Clb5,6		**C**yc**L**in **B** 5,6
Cln1,2		**C**yc**L**i**N** 1,2
DNA polymerase		DNA polymerase ε and δ
Dpb11		**D**NA **P**olymerase **B** (II)
GINS	Psf1-3 and Sld5	**G**o, **I**chi, **N**i and **S**an complex (japanese for 1,2,3,5)
Mcm2-7		**M**ini**C**hromosome **M**aintenance 2-7
Orc1-6		**O**rigin **R**ecognition **C**omplex 1-6
SCF		**S**kp1–**C**dc53/Cull–**F**-box protein complex
Sld2	Drc1	**S**ynthetically **L**ethal with **D**pb11-1 2
Sld3		**S**ynthetically **L**ethal with **D**pb11-1

Table .4: List of abbreviations used in this work

Abbreviation	**Explanation**
ARS	**A**utonomously **R**eplicating **R**equence
ATP	**A**denosin**T**ri**P**hosphat
BrdU	**B**romodeoxy**U**ridine
Cdk	**C**yclin-**D**ependent **K**inase
DDE	**D**elay **D**ifferential **E**quation
FACS	**F**luorescence **A**ctivated **C**ell **S**orting
G1, G2 phase	**G**ap phase 1,2
HU	**H**ydroxy**U**rea
M phase	**M**itotic phase
NES	**N**uclear **E**xport **S**ignal
NLS	**N**uclear **L**ocalization **S**ignal
ODE	**O**rdinary **D**ifferential **E**quation
ORC	**O**rigin **R**ecognition **C**omplex
pre-IC	pre-**I**nitiation **C**omplex
pre-RC	pre-**R**eplication **C**omplex
RC	**R**eplication **C**omplex
S phase	**S**ynthesis phase

Bibliography

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, New York / Abingdon, 2009.

[2] E. E. Arias and J. C. Walter. Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. *Genes Dev*, 21(5):497–518, 2007.

[3] P. Atkins and J. De Paula. *Physical Chemistry*. Oxford University Press, Oxford, UK, 8th edition, 2006.

[4] G. B. and R. P. Total process optimization in chemical engineering with evolutionary algorithms. *Computers & Chemical Engeneering*, 22(1):S229–S236, 1998.

[5] M. Barberis, L. De Gioia, M. Ruzzene, S. Sarno, P. Coccetti, P. Fantucci, M. Vanoni, and L. Alberghina. The yeast cyclin-dependent kinase inhibitor Sic1 and mammalian p27Kip1 are functional homologues with a structurally conserved inhibitory domain. *Biochem J*, 387(Pt 3):639–647, 2005.

[6] M. Barberis, E. Klipp, M. Vanoni, and L. Alberghina. Cell size at S phase initiation: an emergent property of the G1/S network. *PLoS Comput Biol*, 3(4):e64, 2007.

[7] J. Bloom and F. R. Cross. Novel role for Cdc14 sequestration: Cdc14 dephosphorylates factors that promote DNA replication. *Mol Cell Biol*, 27(3):842–853, 2007.

[8] J. Bloom and F. R. Cross. Multiple levels of cyclin specificity in cell-cycle control. *Nat Rev Mol Cell Biol*, 8(2):149–160, 2007.

[9] J. J. Blow and A. Dutta. Preventing re-replication of chromosomal DNA. *Nat Rev Mol Cell Biol*, 6(6):476–486, 2005.

[10] M. Botchan. Cell biology: a switch for S phase. *Nature*, 445(7125):272–274, 2007.

[11] B. J. Brewer, E. Chlebowicz-Sledziewska, and W. L. Fangman. Cell cycle phases in the unequal mother/daughter cell cycles of Saccharomyces cerevisiae. *Mol Cell Biol*, 4(11): 2529–2531, 1984.

[12] K. C. Chen, A. Csikasz-Nagy, B. Gyorffy, J. Val, B. Novak, and J. J. Tyson. Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol Biol Cell*, 11(1):369–391, 2000.

[13] K. C. Chen, L. Calzone, A. Csikasz-Nagy, F. R. Cross, B. Novak, and J. J. Tyson. Integrative analysis of cell cycle control in budding yeast. *Mol Biol Cell*, 15(8):3841–3862, 2004.

[14] S. Chen, M. A. de Vries, and S. P. Bell. Orc6 is required for dynamic recruitment of Cdt1 during repeated Mcm2-7 loading. *Genes Dev*, 21(22):2897–2907, 2007.

[15] K. R. Christie, S. Weng, R. Balakrishnan, M. C. Costanzo, K. Dolinski, S. S. Dwight, S. R. Engel, B. Feierbach, D. G. Fisk, J. E. Hirschman, E. L. Hong, L. Issel-Tarver, R. Nash, A. Sethuraman, B. Starr, C. L. Theesfeld, R. Andrada, G. Binkley, Q. Dong, C. Lane, M. Schroeder, D. Botstein, and J. M. Cherry. Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from Saccharomyces cerevisiae and related sequences from other organisms. *Nucleic Acids Res*, 32(Database issue):D311–4, 2004.

[16] F. R. Cross, V. Archambault, M. Miller, and M. Klovstad. Testing a mathematical model of the yeast cell cycle. *Mol Biol Cell*, 13(1):52–70, 2002.

[17] F. R. Cross, L. Schroeder, and J. M. Bean. Phosphorylation of the Sic1 inhibitor of B-type cyclins in Saccharomyces cerevisiae is not essential but contributes to cell cycle robustness. *Genetics*, 176(3):1541–1555, 2007.

[18] A. Csikasz-Nagy. Computational systems biology of the cell cycle. *Brief Bioinform*, 10 (4):424–434, 2009.

[19] T. Degenhardt, K. N. Rybakova, A. Tomaszewska, M. J. Mone, H. V. Westerhoff, F. J. Bruggeman, and C. Carlberg. Population-level transcription cycles derive from stochastic timing of single-cell transcription. *Cell*, 138(3):489–501, 2009.

[20] R. J. Deshaies and J. E. J. Ferrell. Multisite phosphorylation and the countdown to S phase. *Cell*, 107(7):819–822, 2001.

[21] S. Di Talia, J. M. Skotheim, J. M. Bean, E. D. Siggia, and F. R. Cross. The effects of molecular noise and size control on variability in the budding yeast cell cycle. *Nature*, 448 (7156):947–951, 2007.

[22] J. F. X. Diffley. Regulation of early events in chromosome replication. *Curr Biol*, 14(18): R778–86, 2004.

[23] L. S. Drury, G. Perkins, and J. F. Diffley. The cyclin-dependent kinase Cdc28p regulates distinct modes of Cdc6p proteolysis during the budding yeast cell cycle. *Curr Biol*, 10(5): 231–240, 2000.

[24] W. L. Fangman and B. J. Brewer. A question of time: replication origins of eukaryotic chromosomes. *Cell*, 71(3):363–366, 1992.

[25] L. Feng, B. Wang, B. Driscoll, and A. Jong. Identification and characterization of Saccharomyces cerevisiae Cdc6 DNA-binding properties. *Mol Biol Cell*, 11(5):1673–1685, 2000.

[26] R. R. Gabdoulline and R. C. Wade. Biomolecular diffusional association. *Curr Opin Struct Biol*, 12(2):204–213, 2002.

[27] S. Ghaemmaghami, W.-K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea, and J. S. Weissman. Global analysis of protein expression in yeast. *Nature*, 425 (6959):737–741, 2003.

[28] D. M. Gilbert. Nuclear position leaves its mark on replication timing. *J Cell Biol*, 152(2): F11–5, 2001.

[29] A. Goldar, H. Labit, K. Marheineke, and O. Hyrien. A dynamic stochastic model for DNA replication initiation in early embryos. *PLoS One*, 3(8):e2919, 2008.

[30] A. Goldar, M.-C. Marsolier-Kergoat, and O. Hyrien. Universal temporal profile of replication origin activation in eukaryotes. *PLoS One*, 4(6):e5899, 2009.

[31] B. M. Green, R. J. Morreale, B. Ozaydin, J. L. Derisi, and J. J. Li. Genome-wide mapping of DNA synthesis in Saccharomyces cerevisiae reveals that mechanisms preventing reinitiation of DNA replication are not redundant. *Mol Biol Cell*, 17(5):2401–2414, 2006.

[32] B. Hao, S. Oehlmann, M. E. Sowa, J. W. Harper, and N. P. Pavletich. Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Mol Cell*, 26(1):131–143, 2007.

[33] R. Heinrich and S. Schuster. *The Regulation of Cellular Systems*. Chapman & Hall, New York, NY, 1996.

[34] R. Heinrich, S. Schuster, and H. G. Holzhutter. Mathematical analysis of enzymic reaction systems using optimization principles. *Eur J Biochem*, 201(1):1–21, 1991.

[35] P. Heun, T. Laroche, M. K. Raghuraman, and S. M. Gasser. The positioning and dynamics of origins of replication in the budding yeast nucleus. *J Cell Biol*, 152(2):385–400, 2001.

[36] S. Honey and B. Futcher. Roles of the CDK phosphorylation sites of yeast Cdc6 in chromatin binding and rereplication. *Mol Biol Cell*, 18(4):1324–1336, 2007.

[37] A. E. Ikui, V. Archambault, B. J. Drapkin, V. Campbell, and F. R. Cross. Cyclin and cyclin-dependent kinase substrate requirements for preventing rereplication reveal the need for concomitant activation and inhibition. *Genetics*, 175(3):1011–1022, 2007.

[38] P. Jorgensen, N. P. Edgington, B. L. Schneider, I. Rupes, M. Tyers, and B. Futcher. The size of the nucleus increases as yeast cells grow. *Mol Biol Cell*, 18(9):3523–3532, 2007.

[39] S. Jun and J. Bechhoefer. Nucleation and growth in one dimension. II. Application to DNA replication kinetics. *Phys Rev E Stat Nonlin Soft Matter Phys*, 71(1 Pt 1):011909, 2005.

[40] S. Jun, H. Zhang, and J. Bechhoefer. Nucleation and growth in one dimension. I. The generalized Kolmogorov-Johnson-Mehl-Avrami model. *Phys Rev E Stat Nonlin Soft Matter Phys*, 71(1 Pt 1):011908, 2005.

[41] M. Kanemaki and K. Labib. Distinct roles for Sld3 and GINS during establishment and progression of eukaryotic DNA replication forks. *EMBO J*, 25(8):1753–1763, 2006.

[42] S. Kar, W. T. Baumann, M. R. Paul, and J. J. Tyson. Exploring the roles of noise in the eukaryotic cell cycle. *Proc Natl Acad Sci U S A*, 106(16):6471–6476, 2009.

[43] S. Krishnamurthy, E. Smith, D. Krakauer, and W. Fontana. The stochastic behavior of a molecular switching circuit with feedback. *Biol Direct*, 2:13, 2007.

[44] K. Labib and B. Hodgson. Replication fork barriers: pausing for a break or stalling for time? *EMBO Rep*, 8(4):346–353, 2007.

[45] A. Lengronne and E. Schwob. The yeast CDK inhibitor Sic1 prevents genomic instability by promoting replication origin licensing in late G(1). *Mol Cell*, 9(5):1067–1078, 2002.

[46] A. Lengronne, P. Pasero, A. Bensimon, and E. Schwob. Monitoring S phase progression globally and locally using BrdU incorporation in TK(+) yeast strains. *Nucleic Acids Res*, 29(7):1433–1442, 2001.

[47] E. Li. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet*, 3(9):662–673, 2002.

[48] C. Liang and B. Stillman. Persistent initiation of DNA replication and chromatin-bound MCM proteins during the cell cycle in cdc6 mutants. *Genes Dev*, 11(24):3375–3386, 1997.

[49] M. E. Liku, V. Q. Nguyen, A. W. Rosales, K. Irie, and J. J. Li. CDK phosphorylation of a novel NLS-NES module distributed between two subunits of the Mcm2-7 complex prevents chromosomal rereplication. *Mol Biol Cell*, 16(10):5026–5039, 2005.

[50] D. J. Lockhart, H. Dong, M. C. Byrne, M. T. Follettie, M. V. Gallo, M. S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Horton, and E. L. Brown. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol*, 14(13): 1675–1680, 1996.

[51] S. Lopez-Aviles, O. Kapuy, B. Novak, and F. Uhlmann. Irreversibility of mitotic exit is the consequence of systems-level feedback. *Nature*, 459(7246):592–595, 2009.

[52] M. S. Luijsterburg, G. von Bornstaedt, A. M. Gourdin, A. Z. Politi, M. J. Mone, D. O. Warmerdam, J. Goedhart, W. Vermeulen, R. van Driel, and T. Hofer. Stochastic and reversible assembly of a multiprotein DNA repair complex ensures accurate target site recognition and efficient repair. *J Cell Biol*, 189(3):445–463, 2010.

[53] J. Lygeros, K. Koutroumpas, S. Dimopoulos, I. Legouras, P. Kouretas, C. Heichinger, P. Nurse, and Z. Lygerou. Stochastic hybrid modeling of DNA replication across a complete genome. *Proc Natl Acad Sci U S A*, 105(34):12295–12300, 2008.

[54] H. Masumoto, S. Muramatsu, Y. Kamimura, and H. Araki. S-Cdk-dependent phosphorylation of Sld2 essential for chromosomal DNA replication in budding yeast. *Nature*, 415 (6872):651–655, 2002.

[55] L. Matsson. DNA replication and cell cycle progression regulated by long range interactions between protein complexes bound to DNA. *Journal of Biological Physics*, 27: 329–359, 2001.

[56] M. Mechali. DNA replication origins: from sequence specificity to epigenetics. *Nat Rev Genet*, 2(8):640–645, 2001.

[57] C. G. Moles, P. Mendes, and J. R. Banga. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res*, 13(11):2467–2474, 2003.

[58] P. Nash, X. Tang, S. Orlicky, Q. Chen, F. B. Gertler, M. D. Mendenhall, F. Sicheri, T. Pawson, and M. Tyers. Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature*, 414(6863):514–521, 2001.

[59] V. Q. Nguyen, C. Co, and J. J. Li. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature*, 411(6841):1068–1073, 2001.

[60] C. A. Nieduszynski, S.-i. Hiraga, P. Ak, C. J. Benham, and A. D. Donaldson. OriDB: a DNA replication origin database. *Nucleic Acids Res*, 35(Database issue):D40–6, 2007.

[61] S. H. Northrup and H. P. Erickson. Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc Natl Acad Sci U S A*, 89(8):3338–3342, 1992.

[62] R. Nougarede, F. Della Seta, P. Zarzov, and E. Schwob. Hierarchy of S-phase-promoting factors: yeast Dbf4-Cdc7 kinase requires prior S-phase cyclin-dependent kinase activation. *Mol Cell Biol*, 20(11):3795–3806, 2000.

[63] B. Novak and J. J. Tyson. Modeling the control of DNA replication in fission yeast. *Proc Natl Acad Sci U S A*, 94(17):9147–9152, 1997.

[64] R. Phillips, J. Kondev, and J. Theriot. *Physical Biology of the Cell*. Garland Science, 2009.

[65] A. Politi, M. J. Mone, A. B. Houtsmuller, D. Hoogstraten, W. Vermeulen, R. Heinrich, and R. van Driel. Mathematical modeling of nucleotide excision repair reveals efficiency of sequential assembly strategies. *Mol Cell*, 19(5):679–690, 2005.

[66] D. M. PRESCOTT. Relation between cell growth and cell division. III. Changes in nuclear volume and growth rate and prevention of cell division in Amoeba proteus resulting from cytoplasmic amputations. *Exp Cell Res*, 11(1):94–98, 1956.

[67] M. K. Raghuraman, E. A. Winzeler, D. Collingwood, S. Hunt, L. Wodicka, A. Conway, D. J. Lockhart, R. W. Davis, B. J. Brewer, and W. L. Fangman. Replication dynamics of the yeast genome. *Science*, 294(5540):115–121, 2001.

[68] J. C. W. Randell, J. L. Bowers, H. K. Rodriguez, and S. P. Bell. Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. *Mol Cell*, 21(1):29–39, 2006.

[69] R. L. Rossi, V. Zinzalla, A. Mastriani, M. Vanoni, and L. Alberghina. Subcellular localization of the cyclin dependent kinase inhibitor Sic1 is modulated by the carbon source in budding yeast. *Cell Cycle*, 4(12):1798–1807, 2005.

[70] C. Salazar and T. Hofer. Versatile regulation of multisite protein phosphorylation by the order of phosphate processing and protein-protein interactions. *FEBS J*, 274(4):1046–1061, 2007.

[71] C. Santocanale and J. F. Diffley. ORC- and Cdc6-dependent complexes at active and inactive chromosomal replication origins in Saccharomyces cerevisiae. *EMBO J*, 15(23): 6671–6679, 1996.

[72] Y.-J. Sheu and B. Stillman. Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. *Mol Cell*, 24(1):101–113, 2006.

[73] J. M. Sidorova and L. L. Breeden. Precocious G1/S transitions and genomic instability: the origin connection. *Mutat Res*, 532(1-2):5–19, 2003.

[74] J. M. Skotheim, S. Di Talia, E. D. Siggia, and F. R. Cross. Positive feedback of G1 cyclins ensures coherent cell cycle entry. *Nature*, 454(7202):291–296, 2008.

[75] T. W. Spiesser, E. Klipp, and M. Barberis. A model for the spatiotemporal organization of DNA replication in Saccharomyces cerevisiae. *Mol Genet Genomics*, 282(1):25–35, 2009.

[76] J. Srividhya and M. S. Gopinathan. A simple time delay model for eukaryotic cell cycle. *J Theor Biol*, 241(3):617–627, 2006.

[77] Y.-S. Tak, Y. Tanaka, S. Endo, Y. Kamimura, and H. Araki. A CDK-catalysed regulatory phosphorylation for formation of the DNA replication complex Sld2-Dpb11. *EMBO J*, 25 (9):1987–1996, 2006.

[78] S. Tanaka and J. F. X. Diffley. Deregulated G1-cyclin expression induces genomic instability by preventing efficient pre-RC formation. *Genes Dev*, 16(20):2639–2649, 2002.

[79] S. Tanaka and J. F. X. Diffley. Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2-7 during G1 phase. *Nat Cell Biol*, 4(3):198–207, 2002.

[80] S. Tanaka, Y.-S. Tak, and H. Araki. The role of CDK in the initiation step of DNA replication in eukaryotes. *Cell Div*, 2:16, 2007.

[81] S. Tanaka, T. Umemori, K. Hirai, S. Muramatsu, Y. Kamimura, and H. Araki. CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature*, 445(7125):328–332, 2007.

[82] R. E. Tanny, D. M. MacAlpine, H. G. Blitzblau, and S. P. Bell. Genome-wide analysis of re-replication reveals inhibitory controls that target multiple stages of replication initiation. *Mol Biol Cell*, 17(5):2415–2423, 2006.

[83] R. Visintin, K. Craig, E. S. Hwang, S. Prinz, M. Tyers, and A. Amon. The phosphatase Cdc14 triggers mitotic exit by reversal of Cdk-dependent phosphorylation. *Mol Cell*, 2(6): 709–718, 1998.

[84] G. M. Wilmes, V. Archambault, R. J. Austin, M. D. Jacobson, S. P. Bell, and F. R. Cross. Interaction of the S-phase cyclin Clb5 with an "RXL" docking sequence in the initiator protein Orc6 provides an origin-localized replication control switch. *Genes Dev*, 18(9): 981–991, 2004.

[85] E. Wintersberger. Why is there late replication? *Chromosoma*, 109(5):300–307, 2000.

[86] Wolfram Research, Inc. Mathematica version 7.0, 2009.

[87] P.-Y. J. Wu and P. Nurse. Establishing the program of origin firing during S phase in fission Yeast. *Cell*, 136(5):852–864, 2009.

[88] N. Yabuki, H. Terashima, and K. Kitada. Mapping of early firing origins on a replication profile of budding yeast. *Genes Cells*, 7(8):781–789, 2002.

[89] P. Zegerman and J. F. X. Diffley. Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature*, 445(7125):281–285, 2007.

[90] P. Zegerman and J. F. X. Diffley. DNA replication as a target of the DNA damage checkpoint. *DNA Repair (Amst)*, 8(9):1077–1088, 2009.

Selbständigkeitserklärung

Hiermit versichere ich, die vorliegende Arbeit selbständig und ausschließlich unter Verwendung der angegebenen Mittel und ohne unerlaubte Hilfen angefertigt zu haben.

Heidelberg, den 09.12.2009 Anneke Brümmer